

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 15:48:35 ; Search time 6225.59 Seconds
(without alignments)
14215.251 Million cell updates/sec

Title: US-09-316-163-1
Perfect score: 4229
Sequence: 1 tcgagtcgaactgtcccgag.....aaaaaaaaaaaaaaaaaaaaa 4229

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_inv:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	4229	100.0	4229	6	A91883	A91883 Sequence 1
2	4025.4	95.2	4256	10	RNO320522	AJ320522 Rattus no
3	3172.4	75.0	4300	10	MUSH	M12660 Mouse Cfh 1
4	2360	55.8	2715	6	A91885	A91885 Sequence 3
5	2055.2	48.6	3926	9	HSB	Y00716 Human mRNA
6	1616	38.2	2955	10	MUSCFHRD	M29009 Mouse compl
7	1229.4	29.1	2729	10	AF436847	AF436847 Rattus no
8	1119	26.5	1532	6	A91886	A91886 Sequence 4
9	1016.4	24.0	2132	6	HUMCHF	M17517 Human compl
10	998.4	23.6	2008	4	BTOMPH	X98697 B.taurus mr
11	824.8	19.5	1658	9	HSHT	X07523 Human mRNA
12	693.4	16.4	1427	9	HSB38	X04697 Human mRNA
13	691	16.3	866	6	A91884	A91884 Sequence 2
14	594.6	14.1	2356	10	MUSCFHRC	M29010 Mouse compl
15	493.2	11.7	1486	9	BC012610	BC012610 Homo sapi
16	445.8	10.5	470	10	AF348145S1	AF348145 Rattus no
17	442.6	10.5	1192	10	MUSCFHRA	M29007 Mouse compl
18	426.2	10.1	2823	6	AX274975	AX274975 Sequence
19	426.2	10.1	2823	9	AF295327	AF295327 Homo sapi
20	408.2	9.7	1707	6	AX274977	AX274977 Sequence
21	373.6	8.8	907	9	HUMCHA	M12383 Human compl
22	371.4	8.8	151263	2	AC096417	AC096417 Rattus no
23	368.8	8.7	659	10	AF348145S2	AF348144 Rattus no
24	331.8	7.8	1734	10	MUSCFHRB	M29008 Mouse compl
25	330	7.8	1264	9	HUMHAAB	M65293 Human factio
26	326.8	7.7	1296	9	BC016755	BC016755 Homo sapi
27	326	7.7	1148	9	HSB361	X56209 Human mRNA
28	326	7.7	1196	9	HSCMRH362	X56210 H.sapiens m
29	322	7.6	1266	9	HUMHAAAC	M65292 Human factio
30	319.6	7.6	755	9	HUMHAAAC	M65294 Human factio
31	300.6	7.1	1269	9	HSDOWN	X68679 H. sapiens
32	268.2	6.3	1490	9	BC020687	BC020687 Homo sapi
33	228	5.4	658	10	MUSMFACH2	M23554 Mouse compl
34	208	4.9	435	10	MUSH2A2	M31980 Mouse factio
35	207.6	4.9	151263	2	AC096417	AC096417 Rattus no
36	196	4.6	1315	9	HSCFHRPT4	X98337 H.sapiens m
37	192.2	4.5	1075	9	BC022283	BC022283 Homo sapi
38	190.6	4.5	1040	9	HSDESKS9	X64877 H.sapiens m
39	154.2	3.6	189282	2	AC097078	AC097078 Rattus no
40	151.4	3.6	37033	2	AC100698	AC100698 Mus muscu
41	149.4	3.5	189282	2	AC097078	AC097078 Rattus no
42	146	3.5	37033	2	AC100698	AC100698 Mus muscu
43	132.6	3.1	150626	9	HSJ177P10	AL049744 Human DNA
44	132.6	3.1	180155	2	AL161735	AL161735 Homo sapi
45	131	3.1	186760	2	AC026056	AC026056 Homo sapi

ALIGNMENTS

RESULT 1

A91883	A91883	Sequence 1 from Patent WO9823638.	4229 bp	DNA	linear	PAT 22-JAN-2000
LOCUS	A91883	Sequence 1 from Patent WO9823638.	4229 bp	DNA	linear	PAT 22-JAN-2000
DEFINITION	A91883	Sequence 1 from Patent WO9823638.	4229 bp	DNA	linear	PAT 22-JAN-2000
ACCESSION	A91883	Sequence 1 from Patent WO9823638.	4229 bp	DNA	linear	PAT 22-JAN-2000
VERSION	A91883.1	GI:6740760	4229 bp	DNA	linear	PAT 22-JAN-2000
KEYWORDS	A91883.1	GI:6740760	4229 bp	DNA	linear	PAT 22-JAN-2000
SOURCE	unidentified.	unclassified.	4229 bp	DNA	linear	PAT 22-JAN-2000
ORGANISM	unclassified.	unclassified.	4229 bp	DNA	linear	PAT 22-JAN-2000
REFERENCE	1 (bases 1 to 4229)	unclassified.	4229 bp	DNA	linear	PAT 22-JAN-2000
AUTHORS	Schwaebler, W. and Slin, R. B.	unclassified.	4229 bp	DNA	linear	PAT 22-JAN-2000
TITLE	COMPLEMENT INHIBITOR	unclassified.	4229 bp	DNA	linear	PAT 22-JAN-2000
JOURNAL	Patent: WO 9823638-A 1 04-JUN-1998;	unclassified.	4229 bp	DNA	linear	PAT 22-JAN-2000
FEATURES	SCHWAEBLER WILHELM (GB); UNIV LEICESTER (GB)	unclassified.	4229 bp	DNA	linear	PAT 22-JAN-2000
SOURCE	Location/Qualifiers	unclassified.	4229 bp	DNA	linear	PAT 22-JAN-2000
BASE COUNT	1395 a	791 c	893 g	1150 t		
ORIGIN	/db_xref="taxon:32644"					

Query Match					100.0% Score 4229; DB 6; Length 4229;				
Best Local Similarity					100.0%; Pred. No. 0;				
Matches 4229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	tcgagtcactgctccagatagatccagacatgagactgctcagcagaatatttggc	60						
Db	1	TCGAGTCAACTGCTCCAGATAGATCCAGACATGAGACTGTCACCAAGAATATTGGC	60						
Qy	61	ttatattatgactgttctgtagcagaagattgttaaaggctccctcccaagagaaaatt	120						
Db	61	TTATATTATGGAAGTCTTTGTGTAGCAGAAGATTGTAAAGGCTCTCTCCCAAGAGAAAATT	120						
Qy	121	cagaatctctcaggttcggtctgacactatattcagaagcactcagcaacct	180						
Db	121	CAGAAATTTCTCAGCTTCGGTCTGAACAACTATATTGAGAGGCACTCAGGCAACCT	180						
Qy	181	acaaatgcgccttgatcaccgaacacttggtactattgtaaaagtgcagaagatggag	240						
Db	181	ACAAATGCCGCCCTGATACCGAACACTTGGTACTATTGTAAGATGATGCAAGATGGAG	240						
Qy	241	aatgggtaccttctaaccatacaagatatgtcggaagaagccatgtgggcatcccgag	300						
Db	241	AATGGGTACCTTCTAACCCATCAAGGATATGTCGGAAGGCCATGTGGGCATCCCGGAG	300						
Qy	301	acacaccccttgggtccctttaggtcggcagttgagatctgaaatttgaatttgcaaaagg	360						
Db	301	ACACACCCCTTTGGGTCCCTTTAGCTGGCAGTTGGATCTGAAATTTGGTGCAAAAGG	360						
Qy	361	ttgtttatcatgtgatgaagggtaccacactattaggtgaaattgattaccgtgaatgtg	420						
Db	361	TTGTTTATACATGTGATCAAGGGTACCAACTATTAGGTGAAATTTGATTCCTGTAATGTG	420						
Qy	421	atgcagatgggtggacaaatgatattccaatattgtaagtgtgaagtgtgaagtgtgca	480						
Db	421	ATGCAGATGGGTGGACCAATGATATTCCAATATGTGAAGTTGTGAAGTGTGCGCAGTGA	480						
Qy	481	cagaactggagaatggagaatgtgagtggtgagcggcagaccagaccaggaattatt	540						
Db	481	CAGAACTGGAATGGAAGAAATGTGAGTGGTGCAGCCGACACGACCGAGGAATATTATT	540						
Qy	541	ttgacaggtgttacgtttgaaatgcactccgcttcaagattgaaaggacagaaagaaa	600						
Db	541	TTGGACAGGTGGTACGCTTTTGAATGCAACTCCGGCTTCAAGATTGAAGCAGAGAAAGAA	600						
Qy	601	tgcactgctataaaatggcctctggagcaatgaaagccacagtggtgaaatttctt	660						
Db	601	TGCAGTCTCATAAATGGCTCTGGAGCAATGAAAGCCACAGTGTGTGAAATTTCTT	660						
Qy	661	gcctgccaccacgagttgaaaatggagatgatatctgaaaccagtttacaagaga	720						
Db	661	GCCTGCCACCACGAGTTGAAATGGAGATGCTATATCTGAACACGATTACAAAGGAGA	720						
Qy	721	atgaaagattccaataataatgtaagaaggttctgtgtacaaagaagaggatgctg	780						
Db	721	ATGAAAGATTCCAATATAAATGTAAAGAGGTTTGTGTACAAAGAAAGAGGGATGCTG	780						
Qy	781	tctgcaggggtcttgatggaatcctcagcctctctgtgaaagaatgacatgttgactc	840						
Db	781	TCTGCAGGGTCTCGGATGGAAATCCTCAGCCCTCTGTGGAAGAAATGACATGTTTGACTC	840						
Qy	841	catatttccaaatgggtatctacacactcacaggtattaaacacagaaattgatgataa	900						
Db	841	CATATTTCCAAATGGTATCTACACACTCACAGGATTAAACACAGAAATGTGATGATGAAA	900						
Qy	901	tcagatatgaatgtaaaaaatggcttctctcctgcaaccccgatcacctgtttcaaaagtga	960						
Db	901	TCAGATATGAATGTAAAAATGGCTTCTATCTCGCAACCCGATCATCCTGTTTCAAAAGTGA	960						
Qy	961	caattactggctggtcctcctcgaagatgtagcttgaacacttgattttccacaat	1020						
Db	961	CAATTACTGGCTGGATCCCTGCTCCAAGATGTAGCTTGAACCTTGTGATTTTCCCAAT	1020						
Qy	1021	tcaaacatggagctctgtattatgaagaagcggagacccctacttcccagtaacctatag	1080						

Db	1021	TCAAAATGGAGCTGTGATTATGAAGAAAGCCGGAGACCTTACTTCCAGTACCTATAG	1080						
Qy	1081	gaaagagtcacagctataactgtgacaacgggtttcacacgcttcacagtcatactggg	1140						
Db	1081	GAAAGGAGTACAGCTATAACTGTGACACAGGGTTTACAACGCTTTCACAGTCACTGGG	1140						
Qy	1141	actacctctgttcacagtaaatgggtggagcctgaagttccatgcctcaggaatgta	1200						
Db	1141	ACTACCTTCTGTTGCACAGTAATGGTGGAGCCTGAAGTTCCTCCGCTCAGGCAATGTA	1200						
Qy	1201	ttttccattatgtgaatatggaatcttcatactctgcaaaagaatatatagaggttc	1260						
Db	1201	TTTTCCATTATGTGGAATATGGAGAAATCTTCATATCGCAAGAGATATATAGAGGTC	1260						
Qy	1261	agctgcacaaagtcagtgctcacagtgctctagctctcccaaatgctcgaagactgttcag	1320						
Db	1261	AGCTGCAAAAGTCCAGTGTACAGTGGCTATAGTCTTCCAATGGTCAAGATACATATT	1320						
Qy	1321	attgtacagagaatggctgggtccctccctcccaaatgctcgaagactgttcag	1380						
Db	1321	ATTGTACAGAAATGGCTGGTCCCTCCCTCCCAAAATGCGTCCGTATCAAGACTTGTTCAG	1380						
Qy	1381	tatcagatatagaaattgaaatgggttttttctcgaatctgattacatactatgtctaa	1440						
Db	1381	TATCAGATATAGAAATGAAATGGGTTTTTCTGAAATCTGATTATACATATGCTCTAA	1440						
Qy	1441	atagaaacacggttatagatgaaacagggatatgttaacaaatcccgagagaatatcag	1500						
Db	1441	ATAGAAAACAGGTTAGATGTAAACAGGATATGTAAACAATACCGGAGAAATATCAG	1500						
Qy	1501	gaataattacttctcgaatggatgggtcacctcgacctctcgaatctgatttaagcttctg	1560						
Db	1501	GAATAATTACTTGTCTTCAAGATGGATGTCACCTCGACCTCATGCTTAAGTCTTGTG	1560						
Qy	1561	atagcctctgatttgagaattctcagtaagaataataaacacatcgggtttaaactcaatg	1620						
Db	1561	ATATGCCCTGATTGAGAAATCTATGACTAAGAAATTAACACATGGTGTAACTCAATG	1620						
Qy	1621	acaaattagactatgaatgtcacattggatatgaaatgaataataaacatcaccaagct	1680						
Db	1621	ACAAATTAGACTATGAATGTACATTGGATATGAAATGAATATATAACATACCAAGGCT	1680						
Qy	1681	ctataacatgactatgatggatgggtctagtaacacctcctctgtatgaaagagaatgca	1740						
Db	1681	CTATAACATGACTATTATGATGGTGTAGTACACCTCTCTGTTATGAAAGAGAATGCA	1740						
Qy	1741	gcattccctcttacacacacacacttagttgtttccacagagaagtaaaatacaaatgtg	1800						
Db	1741	GCAITCCCTCTTACACCAAGACTTAGTTGTTTTTCCAGAGAGTAATAATACAAAGTTG	1800						
Qy	1801	gagattcgttgagttctctcgtccgttcagacacagagttggagcagatttagtgcatt	1860						
Db	1801	GAGATTGCTTGAGTTTCTCTTGGCTTCAGACACAGAGTTGGAGCAGATTAGTGGCAAT	1860						
Qy	1861	gctaccactttggatgggtcccttaatttcccaacgtgtgaaggccaaagtaaaatcatgtg	1920						
Db	1861	GCTACCACCTTTGGATGGTCCCTTAATTTCCCAACGTGTGGAAGGCCAAGTAAATCATGTG	1920						
Qy	1921	accacactcttgaatccccgaatggggaataaaagggaacaaaaaagtgtgaatacagcc	1980						
Db	1921	ACCAACCTCTTGAATCCCGAATCCGGAATGGGAAATTAAGGGGAACAAAAGTTGAATACAGCC	1980						
Qy	1981	atggtgacgtggtggaataatgattgcaaaccttagattctactgaaggagcccaataaaa	2040						
Db	1981	ATGGTGAGGTGGTGGATATGATTGCAAACTAGATTCTACTGAAGGGACCCCAATATAA	2040						
Qy	2041	tccagttgttgacgggaagtggacaaggttgcgatatgcgttgatgagagaacat	2100						
Db	2041	TCCAGTGTGTGACGGGAAGTGGACAAGGTGCGGATATGCGTTTGTAGTATGAGAGAAT	2100						
Qy	2101	gtggagacccctctcgaactgagcatggctctgcaagttatctgtccctccctaccatc	2160						

Db 2101 GTGGAGACCTTCTCTGAACCTTGAGCATGCTCTGTCAGTTATATCTGTCTCCCTCCCTACCATC 2160
QY 2161 atggagattcagtgagggttcacttgTaccagaaacccTtcaaatgattgacatgagtag 2220
Db 2161 ATGGAGATTTCAGTGGAGTTCACCTTGTACAGAAACCTTCACAATGATTGCACATGCGAGTAG 2220
QY 2221 ttttctgcattagtgaaggtgagccagacttccctaattgtgttgcaacagatcaactg 2280
Db 2221 TTTTCTGCAATTAGTGAAGGTGAGCCGAGCTTCCTCAATGTGTGCAACAGATCAACTUG 2280
QY 2281 agaagtgtaaagcccggaagtccaactggcatagatgacaattccatccaaataaauaatgaat 2340
Db 2281 ACAAGTGTAAAGCCCGGAAGTCAACTGGCATAGATCCAATTCATCCAANTAGAATGAAT 2340
QY 2341 ttaatacaactttagtgtgagtacagatgtagacaaagcaggagtatgaacattcaa 2400
Db 2341 TTAATCATAACTTTAGTGTTCAGTTACAGATGTAGACAAAAGCAGGAGTATGAACATTCAA 2400
QY 2401 tcltgcataaatggaagatggatccctgaaccaaactgtcaagcaaaaatcttgcacclc 2460
Db 2401 TCTGCGATCAATGGAAGTGGGATCTTGAACCAAACTGTACAAGCAAAAGATTCTGCCCTC 2460
QY 2461 clccccgcagattccaaatgccaaagtgtatgaaaccaccgtgaaatacttggatggag 2520
Db 2461 CTCCCCCGCAGATTCCAAATGCCCAAGTGATTGAAACCCACCGTGAAATACTTCGGATGGAG 2520
QY 2521 aaaaagtalctgttcttccaaagtgttacctaaactcaggycccaagaagaatgggt 2580
Db 2521 AAAAAATATCTCTGTCTTCTGCCAAGATGGTTAGCTTAACCTCAGGCGCCAGGAAGAACTGCTGT 2580
QY 2581 gtaaacatggaaggTggcagtcgttaccacgctgcacgagaaaaatLccatctLcccaac 2640
Db 2581 GTAAACATGCAAGGTGGCAGTCTTACCACGCTGCACGGAAGAAAATTTCCATGTTCCCAGC 2640
QY 2641 cccctaaaltgaacatggatctataatgctgccaggttccctcagaagagagagatL 2700
Db 2641 CCCCTAAAATGAACATGGCATCTATTAACTGCGCCAGGTCTCAGAAAGAGAGGAGAGATT 2700
QY 2701 taattgagtcceagcagttatgaacaggaactacattcagctatttctgtatagatcgat 2760
Db 2701 TAAATGAGTCCAGCAGTATATGAACAGGAACTACATPTCAGCTATTGCTCTAGAGATGGAT 2760
QY 2761 tcaagatatctgaaagaaatagggtaacnctgcaacalTgagaaaaTggagcttcltgcclc 2820
Db 2761 TCAAGATATCTGAAGAAATAGGGTAACCTGCAACATGGGAAATGGAGCTCTCTGCCCTC 2820
QY 2821 gttgtgtggaataccttgttggaaccccaacttcaattcctcttgatgttctLcatg 2880
Db 2821 GTTGTGTTGGAATACCTTGTGGACGCCACCTTCAATTCTCTTGGTATTTGTTCTCATC 2880
QY 2881 aactagaagttaccaatatggagaggaggttacatacaattgttctgaagactttggaa 2940
Db 2881 AACTAGAAAGTTACCAATATGGAGAGGAGGTTACATACAAATTTCTCTGAAGGCTTTTGGAA 2940
QY 2941 ttgatggaccagcatttataaatgtgtaggaggacagTggtctgaacTccraaatqca 3000
Db 2941 TTGATGGACCACCAATTATTAAATGTGTAGGAGGACAGTGGTCTGAACCTCCCAATGCA 3000
QY 3001 taaaaactgattgtgacaacttgccacaatttgaaattggccaaaccgacagaaaaa 3060
Db 3001 TAAAAACTGATTGTGACAACTTGGCCACATTTGAAATTTGCCAAACCGACAGAAAAAGAAAA 3060
QY 3061 aaaaatcatcaggttcaggagaaagtgacattcagatgtccactcngtatcgaaatgg 3120
Db 3061 AAAAAATCATACAGGTTCAGAGAACAAAGTGACATTTCAGATGTCCAGCTCCGTATCGAATUG 3120
QY 3121 atggctctgacattgttcacatgtgttaataccgaaagtggattggacagccgggtatyrcaag 3180
Db 3121 ATGGCTCTGACATTGTACATGTGTATTACCAAGTGGATTGGACAGCGCGGTATGCAAG 3180
QY 3181 ataattcctgtgtgaatccaccacatgtgtccaaatgtctactactaatacaaggrcaaga 3240
Db 3181 ATAATTCTGTGTGAATCCACCACATGTGCCAAATGCTACTATATACTATAACAAGGCACAAGA 3240

QY 3241 ctaaatatccatctggtgacaaagtacgttatgactgtaataaaccttttgaattatttly 3300
Db 3241 CTAAATATCCATCTGGTGCACAAAGTAGCTTATGACTGTAAATAAACCTTTTGAATTATTG 3300
QY 3301 gggaaagtgaagtgtgtgccaaaaacgggalttggacagaaacccacgaaatgcaaaagalt 3360
Db 3301 GGGAAGTGGAAGTGTGTGCCAAAACGGGATTTGGACAGAACCCACCGAAATGCAAAAGATT 3360
QY 3361 caaacgggaaatLtgggcctccctccacttatgtacaaatggagacatcacctccttltcat 3420
Db 3361 CAACAGGAAATGTGGCGCTCTCCACCTATTGACAATGGAGACATCACCTCTCTGTGCAT 3420
QY 3421 taccagtatatgcaccattatcatcagltgaaatatcaatgccagaactattatclactla 3480
Db 3421 TACCAGTATATGCACCATTTATCATCAGTTCAATATCAATCAATGCCAGAACATATTATCTACTTA 3480
QY 3481 agggaaataagatagtaacalTglaaaaTggaagtgggtctcagcccaaccctgctLac 3540
Db 3481 AGGGAATATAGATAGTAGTAACATGTAGAAATGGAAAGTGGTCTCAGCCACCACCTGCTTAC 3540
QY 3541 atgcattgtgtataccagagaatLtatgtgaaaaacataataatagttctcagatlggaggy 3600
Db 3541 ATGCATGTGTGATACCCAGAAGATATTATGGAAAAACATAATATAGTTCTCAGATGGAGGG 3600
QY 3601 aaaaTgcaagatttattLcccaaLcagggggagaatalTgaaLlcatgtgLaaacctggat 3660
Db 3601 AAAATGCAAGATTTATTTCCAATCAGGGGAGAAATATTGAATTCATGCTGTAAACCTGCAT 3660
QY 3661 atagaaattcagagatcacctLccgtLlctgtacaaagtgcattgaggggtLcaatLcaall 3720
Db 3661 ATAGAAAAATTCAGAGGATCACCTCCCTTTGTTGTTACAAAGTGCATTGAGGGTTCACATCAATT 3720
QY 3721 atcccactgtgtataaaatcgctalLacaattatLtagtaaaccttatgtgtatgagaaatgc 3780
Db 3721 ATCCCACTTGTGTATTAATCGTATACAAATATTATTAGTAACCTTTATGGATGAGAAATGC 3780
QY 3781 acatgtatattactaaLacagltLtgaaLttacatttaaatattgttttagctcatLlccclc 3840
Db 3781 ACATGTATATTACTAATACAGTTTGAATTTACATTTTAAATATTGTTAGCTCATTTTCTCTC 3840
QY 3841 taataagtataaaacttttttatagtgtgttaatcagtaactttacagactgtLlgcc 3900
Db 3841 TAAATAAGTATATAAACTTTTTTATATGTTGGTGTAAATCAGTAACCTTTACAGACTGTTGCC 3900
QY 3901 acaaaagcaagaacattcacattccaaactcctaataccaaatatgatgtgtccaaaggacaaa 3960
Db 3901 ACAAAAGCAAGACATTACATTCAAAACTCTTAATCCAAATATGATATGCCAAGGACAAA 3960
QY 3961 ctatgtcttaagcaagaaaaataaatgttagttcttcaatgtctgttttLlccaggacctll 4020
Db 3961 CTATGTCTTAAGCAAGAAAAATAAATGTTAGTTCTTCAATGCTGTTTTTTATTTCAGGACCTT 4020
QY 4021 tcagatttctctygatacccttttttaggtLctgtgttcagtgagtgaggaagcacactg 4080
Db 4021 TCAGATTTCTTCTGGATACCTTTTGTAGGTTCTGATTTCAGTGCAGTGGGAAGACACACTG 4080
QY 4081 actctgacttcaaaattagttatttacttgcgaatcacattaaacccaaactatcataatca 4140
Db 4081 ACTCTGACTTCAAAATTAGTATTACTTGGCAATACATTTAAACCAACCAACTATCATATAATCA 4140
QY 4141 caaatgtatcacagctaattactgtgtcctacctttgttatcaataaagaatactaaagaag 4200
Db 4141 CAAATGTATACAGCTAAATTACTGTGTCTTACCTTTGTATCAATTAAGAAAAATCTTAACAAAG 4200
QY 4201 ttcttgcttaaaaaaataaaaaa 4229
Db 4201 TTCTTGTCTTAAAAAATAAAAAAATAAAAAA 4229

DEFINITION	Rattus norvegicus mRNA for complement inhibitory factor H (FH gene).		
ACCESSION	AJ320522		
VERSION	AJ320522.1 GI:15485712		
KEYWORDS	complement inhibitory factor H, FH gene.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 4256)		
AUTHORS	Demberg,T., Goetze,O. and Schlaef.G.		
TITLE	Rat complement factor H: molecular cloning, sequencing and expression in tissues and isolated cells		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 4256)		
AUTHORS	Demberg,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-AUG-2001) Demberg T., Immunology, Georg August University of Goettingen, Kreuzberggring 57, 37073 Goettingen, GERMANY		
FEATURES	Location/Qualifiers		
source	1..4256		
	/organism="Rattus norvegicus"		
	/strain="Sprague-Dawley"		
	/db_xref="taxon:10116"		
	/sex="male"		
	/tissue_type="liver"		
	/clone_lib="Lambda Triplex"		
	/dev_stage="ages 10 - 12 weeks"		
	/tissue_lib="CLONTECH rat liver 5' stretch plus cDNA library"		
gene	24..4209		
	/gene="FH"		
CDS	24..3734		
	/gene="FH"		
	/codon_start=1		
	/product="complement inhibitory factor H"		
	/protein_id="CAC87513.1"		
	/db_xref="GI:15485713"		
	/translation="MRLSARIIMLMTVCVVAEDCKGPPPRENSEILLGSWSEOLYSE GTQATYKCRPGYRTLGTIVKYKNCWEVPSNRICRKPCHGPDTPFGSFKLAVGS EFEGKAVVYTDGEGYQLLGEIDYREGDADGNTNDIPICEVVKCLPTELENGRIVSO AAEPOBYFQGVVRFECNSGFKIEGOKEMHCSEGLWSNEKPOCVLSCLPLPVENG DGIYLKPVYKNERFOYCKGQFVYKERGDAVCTGSWNPOPSCEEMTCLTPIYPNGI YTPRIKIRHIDDEIRYCKNGLYPATRSPVSKCTITGWIIPAPRCSIAKPCOPPOPKHR LYEESSRRPYFPVPIGKEYSYICDNGFTTPSQYMDYLRTCTVNGWEVPCLRQCIEH SDIEENGFTSESDYTYALNRKTRYRCKQGVVNTGETSGIITCLIQDWSPPRSCIKS CDMPVFENAMTKNNNTWKNDKLDYECYHGYENYKHTKGSITCTYDQWSPTSCYE RECSTPLLDHDLVVFPRVYKYVGDLSLFSFSCRSHRVGADLVQCYHFGWSNPFTECG QVKSQDQPLETPNGEIKGTKKVYSHGDVVEYDCKPRFLKGNKIOCVDGKWTLPF CVEYERTGDLPELHGSVLSVPPYHHGDSVEFTCTETFTMIGHAVVFCISGRWTEL PQCVATDLEKAKAPKSTGIDAHPNKNENFNHNSVSYRCRQKQYEHISICINRWDP EPNCRNEKRFPPPOIPNAQVIETTVKYLDGKEKVSQVLDQGYLTCCPEEMCKHGR WQSLPRCTEIKPSQPPKIEHGSIKSPRSEERLDIESSYEHGTTFSYVCDGFRIS EENRVTNMGKWSLPCVGLPGCPGPSIPGLIVSHELESYQVEEVYVNCSEBGFID GPATKCVGWSEPPKCIKTDGDLNLTPEIAKPTPEKKKSYRSGQVTRCPRPYRM DGSDIVCVNTKILQVPCQNDSCNPPHPNATILTRHKTKYPSGDKVRYDNKPYE LFWGSGDVPTKFTWEPKPKDSTGCGPPPDINGDITSLPLVPYAPLSVBYQQQ NYLLKGNKIVTICRNKWSQPTCLHACVIPEDIMEKHNIVLWRRENAKIYVSQSENI EFMCKPGYRFRGSPPRPTCKIEGHINPTCV"		
polyA_signal	4194..4199		
	/gene="FH"		
polyA_site	4209		
	/gene="FH"		
BASE COUNT	1423 a	788 c	892 g 1153 t
ORIGIN			
Query Match	95.28;	Score 4025.4;	DB 10; Length 4256;
Best Local Similarity	98.33;	Pred. No. 0;	
Matches 4171; Conservative	0;	Mismatches 46;	Indels 28; Gaps 9;

QY	10	ctgctccagatagatcccaagacatgagactotcaagaagaattatttggctattat	69
DB	1	CTGTCCCATAGATAGATCCCAAGACATGAGACTGTCCACAGAATATTTCGCTTATATTAT	60
QY	70	ggactgtttgttagcagaagatlttaaaggctcctctccaagagagaaattcagaatlc	129
DB	61	GCATGCTTTGTCTAGCAGAAGATTGTAAAGGTCTCTCTCAAGAGAGAAATTCAGAAATTC	120
QY	130	tctcagggttcggtgctgaacactatctcagaaggcactcaggcaacactcaaaatgnc	189
DB	121	TCTCAGGTTCGGTGTGAACAACATATATTCAGAGGGCACTCAGGCAACCTCAAAATGCC	180
QY	190	gcctggataccgaacacttggactattgtaaaagatgcaagaatgagaatgagatg	249
DB	181	GCCTTGGATACCGAACACTTGGTACTATTGTAAGATATCGAAGAAATGAGAAATGGGTAC	240
QY	250	ctctaacccatcaaggatgctcggaagaaggccatgtgggcatcccgagagacacacct	309
DB	241	CTTCTAACCCATCAAGGATATGTCGAAAAGGCCATGTGGCATCCCGAGACACACCTT	300
QY	310	ttgggtcctttaggctggcagttggaattgaaatttgaaatttggtgcaaaaggtgtttata	369
DB	301	TTGGGTCTTTTAGGCTGGCAGTTGGATCTGAATTTGAATTTTGGTCCAAAGGTTGTTTATA	360
QY	370	catgtgaaagggtaccacactattagtgaaattgattaccgtgaatgtgagcagatg	429
DB	361	CATGTGATGAAGGGTACCNACTCTTTAGGTGAATTTGATTACCGTGAATGTGATGCAGATG	420
QY	430	ggtgaccaatgatattccaatattgaaagtgtgaaagtgtgctgcccagtgacagaactg	489
DB	421	GGTGACCAATGATATTCCAATATGTGAAGTTGTGAAGTGTTCAGGTGCTTGCAGTGACAGAATCG	480
QY	490	agaatggaagaattgtgagtggtcagccgaacacagaccagaataattattttggaacag	549
DB	481	AGAAATGGAAGAAATGTGAGTGTGAGCCGCAACACAGCAGCAATATATTTTGGACAGG	540
QY	550	tggtacgctttgaatgcaactccgcttcaagattgaaggacagagaaatgcaactgt	609
DB	541	TGCTACGCTTTGAATGCAACTCCGCTTCAAGATTGAAGCAGAGAAAGAAATGCATGCT	600
QY	610	cataaaatggcctctggagcaatgaaaggccaagtgtgtgaaatllcttccctgcccac	669
DB	601	CAGAAAATGGCCTCTGGAGCAATGAAAGCCACAGTGTGTGAAAATTTCTTCCTCTGCCAC	660
QY	670	cacgagttgaaaatggagatgatatctctgaaccagatttacaagagagaatgaaagat	729
DB	661	CACGAGTTGAAAATGGAGATGATATATCTGAAACCAGTTTACAAAGGAGAAATGAAAGAT	720
QY	730	tccaataaaatgtaagcaaggttttgttacaagaagaaggagatgctgctgcacag	789
DB	721	TCCAATATAAATGTAAGCAAGGTTTTGTGTACAAAAGAAAGAGCGCATGCTGTCACCG	780
QY	790	gtctggatgaactcctcagcttctgtgagaagaatgacatggttgactccatatttc	849
DB	781	GTTCTGGATGAAATCCTCAGCCTTCTGTGAAGAAATGACATGTTTGACTTCCATATATTC	840
QY	850	caaatggttatctacacactcacagattaacacacagaattgatgataatcacagatg	909
DB	841	CAAAATGGTATCTACACACTCACAGATTAAACACAGAAATGATGATGATGAATCAATATG	900
QY	910	aatgtaaaatggtctctcctgcaaccccgatcacctgtttcgaagtgtacaattactg	969
DB	901	AATGTAAAAATGGGTATATCTGCAACCCGATCACCTTTCACCAAGGTGACAAATTAATG	960
QY	970	gctggtcctcctcagaatgtagctgaaaccttgattttcccaaatcaaacatg	1029
DB	961	GCTGGATCCCTGCTCCCAAGATGTAGCTTGAACCTTGTGATTTTCCACAAATTCACCAATG	1020
QY	1030	gacgtctgtattgaagaagccgagacccctacttccccagctaccataggaagaggt	1089
DB	1021	GACGCTGTATTATGAAGAAAGCCGAGACCCCTACTTCCCAGTACCTATAGGAAGAGGT	1080

Db	3238	AATATCCATCTGGTGACAAAGTACCTTATGACTGTAATAAACCTTTTCAATATATTTGGG	3297
Qy	3303	gaagtgaagtgaatgtgccc--aaacagggaatttgacagaaacaccacccgaaatgcgaagatt	3360
Db	3298	GAAGTGAAGTGAATGTGCCAAAAGAGGATTTTGGACAGAACACCCGAAATGCCAAGATT	3357
Qy	3361	caacaggaagaatgaggcctcctccacctatttgacaaatggagagacatcacctcttcttcat	3420
Db	3358	CAACAGGAATATGTGGGCTCTCCACCTATTGACAATGGAGACATCACCTCCCTGTGCAT	3417
Qy	3421	taccagatatacgaccattatcatcagttgtaataatcaatgcccagaactatctacttata	3480
Db	3418	TACCAGTATATGCCACATTAATCATAGTTGAATATCAATGGCAGAACATTAATCTCTACTTA	3477
Qy	3481	aggaataaagatagtaacatgtagaataatggaatggatggctcagccaccacccctgcttac	3540
Db	3478	AGGGAATTAAGATAGTACATGTAGAAATGGAAGTGGTCTCAGGCCACCACTGCTTAC	3537
Qy	3541	atgcagtgtgataccagaagatatattgaaaaacataataatagttctcagaatggaaggg	3600
Db	3538	ATGCATGTGTATACCAAGATATTATGGAAAAACATAATAATGTTCTCAGATGGAGGG	3597
Qy	3601	aaaatgcaagatttatcccaatcaggaggagaattatgaattcattgtgtaaacctggat	3660
Db	3598	AAAATGCAAGATTTATTTCCCAATCAGGGCAGAAATATGAATTCATGTTGAACCTGGAT	3657
Qy	3661	atagaaaaatcagaggatcacctccgttttcgtacaaaatgagcattgagggatcacatcaatt	3720
Db	3658	ATAGAAAATTCAGAGGATCACTCCGTTTCGTACAAAGTCAATGAGGCTCACATCAAT	3717
Qy	3721	atcccactgtgtataaaatcgtcatataaatatttagtaaaacctatgatg-----	3772
Db	3718	ATCCCACTTGTGTATAAAATCGCTATACAAATATTATTAGTAACCTTAATGATGAACCTTTG	3777
Qy	3773	---agaaatgacatgatatattactaatcacagtttgaatttaccattt-aaatattgtta	3828
Db	3778	TTTAGAAATGCAATGATATTAATTAATACAGTTTGAATTTACATTTGAATATGTTTA	3837
Qy	3829	gtcatttctctctaataagatatataaaactttttttatatgtggttaattc-agaactttt	3887
Db	3838	GCTCATTTCTTCTTAATAGTATATAAACTTTTATATATGTTGTTTAATCAAGTAACATTT	3897
Qy	3888	acagactgttgccaaagaacacattatcatcctaaactcctaacc-aaatatgata	3946
Db	3898	ACAGACTGTTGCCAACAAAGCAAGAACATTACATTCAAACTCTTAATCCAAATATGATA	3957
Qy	3947	tgccaaagcaaaactatgctcctgaagcaagaaataaatgtagttctcttcaatgtctgttt	4006
Db	3958	TGCCAAAGCAAACTATGCTTAAGCAAGAAATAAATGTTAGTTCTTCAATGCTCTTTT	4017
Qy	4007	ttattcaggaccttcagattttcttggtatcaccttttgt---taggttctgattcacagtg	4064
Db	4018	TTATTCAAGACTTTTTCAGATTTTCTTGGATACCTTTTGTGTTAGTTCTGATTCACAGTG	4077
Qy	4065	agtggagaacacactgactcgtgactcacaattagttacttacttgcacataacacacca	4124
Db	4078	AGTGGAAAGACACACTGACTCTGACTTCCAAATAGTATTACTTGCCAATACATACACCA	4137
Qy	4125	aactatcataatcacaaatgtatacagctaatattactgtgtctactcttgtatcataa	4184
Db	4138	AACATATCATATACAAATGTATACAGCTAATTAATGTTGTCCTACCTTTGTATCAATA	4197
Qy	4185	aagaaatcctaagaagttctgtctaaataaaaaaaaaaaaaaaaaa 4229	
Db	4198	AAGAAATCTAAGAAAAAATAAAAAAAAAAAAAAAAAAAAAA 4242	
RESULT	3		
MUSH			
LOCUS		4300 bp	linear
DEFINITION		Mouse CFH locus, complement protein H gene, complete cds, clones	
ACCESSION		M12660	

VERSION	M12660.1	GI:193724
KEYWORDS	complement protein H; protein H; serum glycoprotein.	
SOURCE	Mus musculus (strain C57/B10.WR) male liver cDNA to mRNA; Mus musculus (strain C57/B10.WR) male liver DNA; and Mus musculus (strain C57/B10.WR) male liver cDNA to mRNA.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 4300; 1 to 3425; 3474 to 4300)	
TITLE	Kristensen, T. and Tack, B.F.	
JOURNAL	Murine protein H is comprised of 20 repeating units, 61 amino acids in length	
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967 (1986)	
COMMENT	86233353 Draft entry and clean copy sequence for [1]. [Unpublished (1986) Scripps Clinic and Res Found, La Jolla, CA 92037] kindly provided by T. Kristensen, 28-JUL-1986.	
FEATURES	Location/Qualifiers	
source	1. 4300	
	/organism="Mus musculus"	
	/strain="C57/B10.WR"	
	/db_xref="taxon:10090"	
	/sex="male"	
	/tissue_type="liver"	
exon	<1..3425	
	/note="H protein"	
	/number=1	
sig_peptide	101..154	
	/note="H protein signal peptide"	
CDS	Join(101..3425,3474..3853)	
	/note="precursor"	
	/codon_start=1	
	/product="complement protein H"	
	/protein_id="AA33759.1"	
	/db_xref="GI:387181"	
	/translation="MRLSARIILWLTVCAEDCKPPPRENSEILSCSWSEQLYPE GFQATVCRPGVYTLCTIVKCKNGKWAASNPSCRKKPCGHPGDTFGSFLAVGS QFEAGKVVYTCDDGVLLEIDYRECGADGWIINDPLCEVVKCLPVTLENGRIVSG AAETQEVYFGQVRFECNSGFKIEGKHEIHCSENGLWSSOPCEKRCSPYILNCI DGINKVPIKENERHYKCKGKGVKPKRGDVACTGSGWSSOPCEKRCSPYILNCI YTPHRIIHRSDDEIRYECNFGFPTGTSVSKCTPTGWIIPVRCPLKCFEFPQKVFH LYVESLRPNFVSGKYSKVCNDSYSGYSLONGQDPTMTCTENGWSPPPKIRIKTCSA VYVNGDSAYWEKVVYQGSGLKVCNDSYSGYSLONGQDPTMTCTENGWSPPPKIRIKTCSA RCSVPTLDRKLVYSPKRYKRVGDLLEFSCHSGHRVGPDSVOCYHFGWSPGPTCKG QVASCAPPLEILNGLNGEINGAKKVEYSHGEVYKDCRPFLLKGNKIQCVQDGNMTLPV CIEERTCGDIPELEHGSACKSVPYPYHHGDSVEFICEENFTMHGHSVSCISCKWTUL PKCATDQLEKCRVLKSTGTEIAIKPLTEFTHNSMDYKCRDKQOEVSICINGKWDPP EPNCTSKTSCPPPPQIPNTQVIETTVYLDGKLSVLCODNYLTQDSEENYKQDGRMQ SLPRGTEKIPCSOPPTIEHGSINLPRSEERDSIESSSHEHGTTFSYVCDGDFRIPE ENRITCMGKNSPTPRCVGLPCGPPPSIPLGTVSLELESYOHGEEVYHCSGTGFDIG PAPFICEGKNSDPPKCIKTDQVLPVKNAILRGKSKSYRTGKEQVYFRCSQPYOMN GSDTVTCVNSRWITKPCVKNDSQDPPHPVNATIVTRKNKYLHGDRVRYECNKPLEL FGQVEMCENGIIWTEKPCRDSTGCGPPPDINDGQITSLSPVTEPLSSVEYQCKRY YLLGKRTICTNGKNSPEPTCLHACVPIENIMESHNIILKWRHTEKIYSHSGEDIEF CKYGYKARDSPPFKTCINGTINYPVCV"	
mat_peptide	Join(155..3425,3474..3850)	
	/note="H protein"	
	/product="complement protein H"	
intron	3426..3473	
	/note="H cds intron"	
exon	3474..4300	
	/number=2	
BASE COUNT	1401 a 814 c 902 g 1183 t	
ORIGIN	554 bp upstream of XhoI site; chromosome 1.	

Query Match 75.08; Score 3172.4; DB 10; Length 4300;
Best Local Similarity 85.68; Pred. No. 0;
Matches 3622; Conservative 0; Mismatches 551; Indels 0; Gaps 6;

Db	2229	ATGGAGATT	CAGTGGAGTT	TCATTGTTGAAGAAAACCTT	CACAATGATTTGGACATGGGTCAG	2228
Qy	2221	ttttctgcattagtggaagtgtagccagagcttcttccaatgtytltgcacagatcaactcgg				2280
Db	2289	TTTCTTGCAATTACTGCAAAATGGACCCAGCTTCCTTAATGTGTTCCACACAGACCACTCG				2348
Qy	2281	agaagtglaaagcccggaaglcacactgycgatagatgccaatttcattccaaataagaatgaat				2340
Db	2349	AGAAGTGTAGAGTGTGAAGTCAACCTGGCATAGAAGCAATAAAAACCAAAATTTGACTGAAT				2408
Qy	2341	ltaactataactttagtltgagttcacagltgagacaaaacagagagltgaaacattcaa				2400
Db	2409	TTACGCATAACTCCACCATGGATTACAAATGTAGAGACACAGCAGAGTACGAAACGCTCAA				2468
Qy	2401	tcgtcatcaatggaagatgggattcctcgaaacaaactgtacaagcaaaagattcttgccttc				2460
Db	2469	TCTGTATCAATGGAATGGGATCTCTGAACCAAACTGTACAAAGCAAAACATCTCTGCCCTC				2528
Qy	2461	ctccccgcagatltccaaatgccccaaigtgattgaaaccacccgtgaataacttggatugag				2520
Db	2529	CTCCACCGCAGATTCCAAATACCCAAAGTGATTGAACCAACCCGTGAATATCTTGGATGGAG				2588
Qy	2521	aaaaagtattctgttctttccaaagtgttacctaaactcagggtcccagagaagaataatgtgtl				2580
Db	2589	AAAAATTTATCTGTCTTTGCCAAGCAATTAACCTAACCTCAGACTCAGAGAAGATGCTGT				2648
Qy	2581	gtaaacatggaagtgtggaagtgttaccacgctgcacggagaaaaatctcatgttccccagc				2640
Db	2649	GCAAAAGTGAAGGTGGCAGTCTATTACCTTCGCTGCATTTCAAAAATTTCCATGTTTCCGACG				2708
Qy	2641	ccctlaaaattgaacatggatctattaaagtgcgccaggtctctcagaagagagagagatll				2700
Db	2709	CCCTACAATAGAACAATGGATCTATTAAATTTACCACAGATCTTCAGAGAAGAGAGATTT				2768
Qy	2701	taattgagtcacagcttatgaaacacgaactacaattcagctattgtcgttagagatggtat				2760
Db	2769	CCATTTGAGTCCACAGCTCATGAACATGGAACCTACATTCAGACTATGTCTGTGATGATGGTT				2828
Qy	2761	tcaagatattctgaagaaaaatagggtlaacctgcacacatgggaaatggagctctctgccttc				2820
Db	2829	TCAGGATACCTGAAGAAANTAGATAACTCGTACATCGGAAATGAGACACTCCACCCTC				2888
Qy	2821	gttgtgtggaaatcacctctdggacccccacacctccaattctctctgtgtattgtlltctcatg				2880
Db	2889	GCTCTGTGTGGACTTCTCTGTGGACCTCCACCTTCAATTTCTTTGGTACATGTTCTCTTC				2948
Qy	2881	aactlagaagaattaccaatltgagagagaggtttacalacaattgtctctgaagcttllugaa				2940
Db	2949	AGCTAGAGACTTACCACATGGGGAGAGGTTTACATACCATTTGTTTACAGGCTTTGGAA				3008
Qy	2941	ttgatggaccagcatttataatgtgttaggagagagtgctctggaacnctcccaatgna				3000
Db	3009	TTGATGCGACGACATTTATATATGCGAAGGAGGAAAGTGTCTGTGACCCACCAAAATGCA				3068
Qy	3001	taaaactgatgtgacaactgccacatttgaaattggccaaaccgacagagaagaagaaa				3060
Db	3069	TAAAACGGATGTGACGTTTTTACCACAGTTTAAAAATGCCCATTAATTAAGAGGAAAGACA				3128
Qy	3061	aaaaatcatacaggttcaggagacaagtgcattcagatgtccaccctlccgtatctgaatgg				3120
Db	3129	AAAAATCATATAGGACAGGAGAAACAAGTGACATTCAGATGTCAATCTCTCTTATCAATCA				3188
Qy	3121	atggctctgaacattgtcacatgtgtltaatacgaagtggaattggaacacccggttatgcagaag				3180
Db	3189	ATGGCTCAGACACTGTGACATGTGTTAATAGTCGGTGGATTGGACAGCCCAAGTATGCAAG				3248
Qy	3181	ataattctctgttgaattccaccacatgtgccaaaatgctactatactaaacaaggtcaacaga				3240
Db	3249	ATAATTTCTGTGGATCCACCACATGTGCCAAATGCTACTATTAGTAAACAAGGACCAAGA				3308
Qy	3241	ctaaatctccatctggtgacaaaagtacgttatgtactgtactaataaacctlttgaaltatbtg				3300
Db	3309	ATAAATATTCTACATGGTGACAGTACCTTTATGAATGTAAACCCYTTTGAACACTATTTG				3368

RESULT		4			
A91885					
LOCUS					
DEFINITION	A91885	Sequence	2715 bp	DNA	linear
ACCESSION	A91885	from Patent	WO9823638.		PAT 22-JAN-2000
VERSION	A91885.1				
KEYWORDS			GI:6740762		

SOURCE unidentified.
ORGANISM unidentified
REFERENCE 1 (bases 1 to 2715)
AUTHORS Schwaebler, W. and Sim, R. B.
TITLE COMPLEMENT INHIBITOR
JOURNAL Patent: WO 9823638-A 3 04-JUN-1998;
SCHWAEBLER, WILHELM (GB); UNIV LEICESTER (GB)
FEATURES Location/Qualifiers
source 1..2715 /organism="unidentified"
BASE COUNT 863 a 518 c 618 g 716 t
ORIGIN

Query Match 55.88; Score 2360; DB 6; Length 2715;
Best Local Similarity 88.78; Pred. No. 0;
Matches 2715; Conservative 0; Mismatches 0; Indels 345; Gaps 1;

Qy 1 tcgagtcacactgctcccgatagatcccaagacatgagctgctcagcagaattatttggc 60
Db 1 TCGAGTCAACTGCTCCGATAGATCCAGACATGACACTGCTCAGCAGAAATATTGGC 60

Qy 61 ttattatggaactgtttgtgtgagcagaagattgaaagtcctccccaagagaaatt 120
Db 61 TTATATTATGACTGTTGTGTAGCAGAAGATTGTAAGGTCCTCCTCCAAGACAAATT 120

Qy 121 cagaaattctcaggttcggtgctgaacaactatttcagaagcactcagcgaacct 180
Db 121 CAGAAATTCCTCAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180

Qy 181 acaaatgccctgggataccgaacacttggtactattgtaaaagtatgcaagaattggg 240
Db 181 ACAAAATGCCCTGGATACCGAACACTTGGTACTATTGTAAGATATGCAAGATCGAG 240

Qy 241 aatgggtacattctaaacctcaaggtatgctggaagaaagccatgtgggcatcccgag 300
Db 241 AATGGGTACCTTTAAACCCATCAAGGATATGTCGGAAGGCCATGTGGCATCCCGAG 300

Qy 301 acacaccccttgggctctttagctggcagctggtgatctgaatttgaaattgggcaagg 360
Db 301 ACACACCCCTTGGGCTCTTGGCTGGCAGCTGGATCTGAAATTGAAATTGGTGCAAGG 360

Qy 361 ttgtttatcatatgtagaagggtaccacactatttagtgaaattgattaccgtgaattg 420
Db 361 TTGTTTATACATGCTGATGAAGGGTACCACACTATTAGTGAAATTGATTACCGTGAATGTG 420

Qy 421 atcgagatgggtgaccacaaatgatattccaatatgtgaagltgtgaagtgcttgcagly 480
Db 421 ATCGAGATGGGTGACCAATGATATTCCAAATATGTGAAGTTGTGAAGTGTGCCAGTGA 480

Qy 481 cagaactggggaatggaagaattgtgagtggtgagccggaaccagacagcaagaataatt 540
Db 481 CAGAACTGGAGATGAGAGAAATTTGAGTGGTGCAGCCGGAACCCAGACAGGAATATTATT 540

Qy 541 ttggacaggtggtacgcttttgaactccggttcaagattggaagcagacagaaagaa 600
Db 541 TTGGACAGGTGTTACGCTTTGAAATGCAACTCCGCCCTTCAAGATTGAAGGACACAGAAA 600

Qy 601 tgcactgctcaataaattggcctctgagcaaatgaaagccacagtgatgagaatttctt 660
Db 601 TGCACTGCTCAATAAATGGCCCTGAGAGCAATGAAAGGCCACAGTGTGTG- 650

Qy 661 gctgcccacacaggttgaaaaatggagatggatatatctgaaacacagtttacaagaga 720
Db 651 ----- 650

Qy 721 atgaagaattccaataataaattgaagaaggttttgtgtacaagaagaagggatgctg 780
Db 651 ----- 650

Qy 781 tctgcacgggttctggatggaatcctcagccttctctgtgaaagaaatgacatglttgactc 840
Db 781 ----- 840

Db 651 ----- 650

Qy 841 catattccaaatggtatctacacacctcacaggattaaacacagaaatgatgatgaaa 900

Db 651 ----- 650

Qy 901 tcagatatgaatgtaaaaaatggcttctatctcctgcaaccccgatccacctgtttccaaagtga 960

Db 651 ----- 650

Qy 961 caattactggctggatccctgctccaagatgtagcttgaaacctgtgatatttccacaat 1020
Db 651 -----TTGAAACCTTGTGATTTCCACAAT 675

Qy 1021 tcaaacatggagctgtgtattatgaagaaagccggagaccctacttcccagtaacctatag 1080

Db 676 TCAAACTGGAGCTCTGTATTATGAAGAAAGCCGAGACCTACTTCCAGTACCTATAG 735

Qy 1081 gaaaggagtcagctataaactgtgacaacgggtttacaacgccttccagtcatactagg 1140

Db 736 GAAAGGAGTACAGCTATAACTGTGACAAACGGGTTTACAACCCCTTTCACAGTCATACCTGG 795

Qy 1141 actaccttctgtcacagtaaatgggtgggagcctgaagttccatgcctcagggcaatgta 1200

Db 796 ACTACCTTCTGTCACAGTAAATGGGTGGAGCCTGAAGTTCCATGCCCTCAGGCNAATGTA 855

Qy 1201 ttttccattatgtggaatatggagaattcttcactctggcgaagaagatatataagggctc 1260

Db 856 TTTTCCATTATGTGGAATATGGGAATCTTCATACTGSCAAAGAAGATATATAGAGGCTC 915

Qy 1261 agtctcaaaagtcaggtgcacagtgctatagcttcccaaatgcttccaaatggccaatacatatt 1320

Db 916 AGTCTGCAAAAGTCCAGTGTCCAGTGCCTATAGTCTTTCACAAATGGTCAAGATACATATT 975

Qy 1321 attgtacagagaatggctgggtccctcctccccaatgctgcctgcctatcagacttcttcaag 1380

Db 976 ATTGTACAGAGAATGGCTGGTCCCTCCTCCCAAAATGGCTCCGATATCAAGACTTGTTCAG 1035

Qy 1381 tacaagatatagaanaattgaaaaatgggttttttctgtaactgtatatacatatgctctaa 1440

Db 1036 TATCAGATATAGAANAATGAAAATGGGTTTTTTCTGAATCTGATTATACATATGCTCTAA 1095

Qy 1441 atagaaaaacacaggtatagatgtaaacagggatgtaaacataacacagcaagaataacag 1500

Db 1096 ATAGAAAACACGGTATAGATGTAACAGGGATATGTAACAAATACCGGAGAAATATACAG 1155

Qy 1501 gaataataactgtcttccaagatgggtgcacctgcacctcctcatgatttaagttctgtg 1560

Db 1156 GAATAATTAATCTTCAAGATGGTGGTCCCTCAGCCCTCATGCTATTAAGTCTTGTG 1215

Qy 1561 atatgctgtatttgagaattctatgactaagaataataaacacacatggttttaaacctaatg 1620

Db 1216 ATATGCTGTATTGTGAAATTTCTATGACTAAGAATAATAACACATGCTTTTAACTCAATG 1275

Qy 1621 acaaaattagactaatgacatggaatggaataataataaacacatacaaaaagct 1680

Db 1276 ACAAAATTAGACTATGAATGTCACATGGATGGAATGAAATGAATATAAACATACCAAAAGCT 1335

Qy 1681 ctataacatgacttattgtatggatgggtctagtaacacctcctgtttatgaaagagaatgca 1740

Db 1336 CTATAACATGCTATTATGATGGATGGTCTAGTACACCTCCTGTGTATGAAGAGAAATGCA 1395

Qy 1741 gcatccctctgttacaccaagacttagttgttttccagagaaagtaaaatacaaaagt 1800

Db 1396 GCATTCCTCTGTTACACCAAGACTTAGTTGTTTTTCCAGAGAAGTAAATATACAAAGTTG 1455

Qy 1801 ggaattcgttgagtttctctctgctcagacacagaggttgaggacagattttagtgcaat 1860

Db 1456 GAGATTCGTTGAGTTTCTTCTTGGCTTCAGGACACAGAGTTGGAGCAGATTTTGTAGTCAAT 1515

Qy 1861 gctaccacttggatggctccctcaatttcccaacgtgtgaagcccaagtaaaatcatgtg 1920
Db 1861 ----- 1920

```
Db 1516 GCTACCACCTTTGGATGCTCCCTTAATTTCCCAACGTGTGAAGGCCAAGTAATAATCATGTG 1575
Qy 1921 accaacctcttgaaatcccgaaatgggaaataaaagggaacaaagaaagttgaataacagcc 1980
Db 1576 ACCAACCTCTTGAANTCCGGAATGGGCAATAAAGAGGAACAAAAAACTTGAATACAGCC 1635
Qy 1981 atggtgacgtggtggaatatgatgcaaacctagattcttactgaaaggacccaataaaa 2040
Db 1636 ATGCTGACCTGGTGAATATGATTGCAAAACCTAGATTCTACTCAAGGGACCCCAATAAAA 1695
Qy 2041 tcaagtgttgacggaagtggacaagttggcgatatcgatcggttgagtatgagagaacat 2100
Db 1696 TCCAGTGTGTGACGGGAAGTGGACAAGTTGGCGGATATGCGTGTGAGTATGAGAGAACAT 1755
Qy 2101 gtgagagaccttcctggaacttgagcatggctctgtcaagttatctgtcctccctaccatc 2160
Db 1756 GTGCAGACCTTCCCTGAACCTTGAGCATGGCTCTGTCTGTCAAGTTATCTGTCCCTACCATC 1815
Qy 2161 atgagagattcaagtgttcaacttgatcacagaaccttcacaaatgattggacatgcaglaq 2240
Db 1816 ATGGAGATTTCAGTGGAGTTTCACTTGTACAGAAACCTTCACAATGATTGGACATCGAGTAG 1875
Qy 2221 tttcttcattagtggaaggtggaccagcttctcctcaatgtgttgcaacagatcaactgg 2280
Db 1876 TTTTCTGCATTAGTGAAGGTGGACCGAGCTTCTCTCAATGTGTGCAACAGATCAACTGG 1935
Qy 2281 agaagtglaagcccccgaagtcgaactggcatagatgcaattatcctccaaaataagaatgaat 2340
Db 1936 AGAAGTGTAAAGCCCGCAAGTCAACTGGCATAGATGCAATTTCATCCAAATAAAGAAATGAAT 1995
Qy 2341 ttaatcataacttttagtgaagttacagatgtagacaagaagcaggagatgataacatlcga 2400
Db 1996 TTAATCATATACTTTAGTGTGAGTTACAGATGTAGACAAAACGAGGATATGAACATTTCAA 2055
Qy 2401 tctgcatcaatggaagatgggatactctgaacccaaactgtacaagcaaaagattctgcctc 2460
Db 2056 TCTGCATCAATGGAAGATGGATCTCTCAACCAAACTGTACAAAGCAAAAGATTCTGCCCTC 2115
Qy 2461 ctcccccgagattccaaatgcccgaatgattgaaacccacgtagaaatacttggaatggag 2520
Db 2116 CTCCCCCGCAGATTCCAAATGCCCCAAGTGATTGAACACACCGTGAATACTTGGATGCAG 2175
Qy 2521 aaaaaglatctgttcttgcgaagatggttaccttaactcaggcccgagagaagaatgagtt 2580
Db 2176 AAAAAAGTATCTGTTCTTGGCAAGATGGTTACCTTAACTCAGGCGCCCAAGAAAGATGGTGT 2235
Qy 2581 gtaacacatggaagtgagcagtcgttacacacgctgcacgaggaataatccatgcttcagc 2640
Db 2236 GTAACATGGAAGGTGCGAGTCGTTACCCAGCTGCACGGAAAAAATTTCCATGTTTCCAGC 2295
Qy 2641 cccctaaaattgaacatggaatctattaaagtcgcccaggtcctcagaaagagagagagatt 2700
Db 2296 CCCCCTAAATTTGAACATGGATCTATTAAAGTCGCCCCAGGTCTCTCAGAAAGAGGAGAGATT 2355
Qy 2701 taattgagtcagcagttatgaacacggaactacattcagctattctgctgagagatgagat 2760
Db 2356 TAATTGAGTCCAGCAGTTATGAACACGGAACACTACATTCACTATTCTCTGAGAGATGCAT 2415
Qy 2761 tcaagatatctgaagaaaaatagggttaacctgcacaatggaataatggagctctctgctc 2820
Db 2416 TCAAGATATCTGAAGAAAAATAGGGTAACCTGCAACATGGGAAAAATGGAGCTCTCTGCCCTC 2475
Qy 2821 gttgtgtggaataacttggagcccccccaacttcaattcctcttggtattgtttctcatg 2880
Db 2476 GTTGTGTTGGAATACCTTGTGGACCCCCCACTTCAATTCTCTTGTGATTGTTTCTTCATG 2535
Qy 2881 aactagaaagtaccataatggagagaggttatacatacaattgtcttgaaggcttttggaa 2940
Db 2536 AACTAGAAAGTTTACCAATATGAGAGAGGATTACATACAAATGTTCTGAAGGCTTTGGAA 2595
Qy 2941 ttgatgaccagcattattataatgtgtaggagagcagtggttgaaacctccccaaatgca 3000
Db 2596 TTGATGGACAGCATTTATTAAATGTGTAGGAGGAGAGTGGTCTGAACCTTCCCAATGCA 2655
```

```
Qy 3001 taaaaactgattgtgacaacttcccacatttgaattgaaattgcaaacccagacagaaagaaaa 3060
Db 2656 TAAAAACTGATTGTGACAACATTGCCACATTTCGAATTTGCCAAACCCAGACAAAAAGAAAA 2715
```

RESULT 5

```
SHS
LOCUS HSH 3926 bp mRNA linear PRI 08-APR-1997
DEFINITION Human mRNA for complement factor H.
ACCESSION Y00716 M32093 X07525
VERSION Y00716.1 GI:31964
KEYWORDS complement factor H; glycoprotein; regulatory protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3926)
Day, A. J.
Direct Submission
JOURNAL Submitted (25-FEB-1988) Day A. J., MRC Immunochimistry Unit, Dept.
of Biochemistry, University of Oxford, South Parks Road, Oxford,
OX1 3QU
2 (bases 1 to 3926)
Ripoche, J., Day, A. J., Harris, T. J. and Sim, R. B.
The complete amino acid sequence of human complement factor H
Biochem. J. 249 (2), 593-602 (1988)
88134059
see Immunogenetics 27:211-214(1988) concerning TYR <-> HIS at AA
384
```

FEATURES

```
Location/Qualifiers
1..3926
/organism="Homo sapiens"
/db_xref="taxon.9606"
/clone="B-38-1, R2a and PE3"
/tissue_type="liver"
/clone_lib="human liver cDNA in pti53/PvuII/8"
74..127
/codon_start=1
/product="factor H"
/protein_id="CAA68704.1"
/db_xref="SWISS-PROT:P08603"
/translation="MRLAKIIICLMWALICVAEDCNELPPRRNTEILTGWSDDQTYPE
GTQAIYKCRPGVRSICNVIMVCRKGWALNPLRKQKPCGHPGDTPTGCTTLTGNN
VREYGVKAVYTCNEGYYQLLGEINVRCDTDGWTNDIPICEVVKCLPVTAPENKIVSS
AMEPDREYHFGQAVRFVCSGYKIEGDEMHCSDDGFNSKEKPKCVSEICKSPDIVNG
YSLPIRLKHTGDEITYOCRNGEYPATRGNTAKCTGTGTPAPRCTLKPCDYPDKHGG
LYENHMRPYPVAVGKYSYDCDEHFETPSGYSWDHIHCTQDQWSPVAPCLKKCYFP
YLENGYNQHGRRKFVQKSIDVACHPGYALPKAOTTVTCTMENGWSPTPRCIRVKTCCK
SSIDJENGFISEOYTYALKKAKYCKLGYVTADGETSGSIRCGKDGWSAOPTCKS
CDIPVFNARTKNDFTWPKLNDTLDECHDGYESNTGSTTGSIVCGYNGWSDLPICYE
RECELPKIDVHLVPPDRKKDOYKGEVLKTSCKPGFTIVGPSVQVHIFGLSPDLPICK
EQVQSGPPPELLNGVYKETEKEYEYSEVYECNPRFLMKGPKNKIQCVDGSMWTLIP
VCIVEESTGCDIPELHGWAQLSSPPYYGDSVEFNSCESFTMIGHRSITC IHGVWTO
LPQCVADIKLKKCKSSNLIILEHLKNKFDHNSNIRYRCRKGWIHTVC INGRWD
PEVNCMAOIQLCPPPPOIPNSHNWTTTLNRYDGEKVSVCQENYLIQEGEELTCKDG
RWQSIPLCVERKIPCSOPPOIEHTINSSRSQESYAHGKLSITCEGGFRISENETT
CYMGKSSPPQCEGLCKSPPEISHGVVAHMSDSYQGEVTKYKCEGFGIDGPAIAK
CLGERKSWPSCITKTDCLSLPSFENAIPMGKRDVYKAGEQVYTKCATYTKMDGASNV
TCINSRWTRPTCTSCVNPPTQVNAVIVSRQMSKYPGSERVYRCRYPYEMFGDEE
VNLGNWTEPPQCKDCTCKCGPPPPIIDNGDITSPFLSVYAPASSVEYQCNLYOLEG
NKRTITRCNGQSEPPKCLHPCVISRIMENYNIARLWTAQKLSTYRSGSEVFCVRG
YRLSSRSHTLRITTCWDGKLEYEPTCAAR"
```

sig_peptide CDS

mat_peptide

```
128..3766
/product="factor H"
994
```

variation

```
/note="polymorphism"
```

variation

```
/replace="c"
1277
```

[illegible]

Db 1966 TCCACCTCTGAACTCCTCAATGGGAATGTTAAGGAAAAAACGAAAGAATATGACACA 2025
QY 1982 tggctgctggtggaataatgattgcaaacctagatttctactgaaggagcccaataaaaa 2041
Db 2026 CACTGAAGTGGTGGAAATATTATTGCAATCCTACATTTCTAATGAAGGACCTAATAAAT 2085
QY 2042 ccaagtgttgacggaagtggacaaggttgccgatatatcggttgagtatgagagaacatg 2101
Db 2086 TCAATGTGTGATGGAGAGTGGACAACTTTACCAGTGTGTATTGTGTGGAGGAGTACCTG 2145
QY 2102 tggagacctctcgaacttgagcatggctctgtcaagttatctgtccctccctaccatca 2161
Db 2146 TGGAGATATACCTGAACCTGCAACATGGCTGGGCCAGCTTTCTCCCTCTTATTACTPA 2205
QY 2162 tggagattcaagtggagttccactgtgtacagaaaccttccacaatgattggacatcgatgt 2221
Db 2206 TGGAGATTCAAGTGGAAATTCAAATGGCTCAGAATCATTTACAATGATTGGACACAGATCAAT 2265
QY 2222 ttctcattagtggaagtggaccagcttctcctaagtgttggaacagatcaactgga 2281
Db 2266 TAGCTGATTTCATGGAGATTGGACCACTTCCCAAGTGTGTGGCAATAGATAAATCTTA 2325
QY 2282 gaagtgaagcccccgaactggaactggcatagatgcaattctccaaataaagaatgaatt 2341
Db 2326 GAAGTCAAAATCATCAATTTAATTTAATCTAGGGAACATTTTAAAAACAAGAAGCAAT 2385
QY 2342 taatcaataacttttagtgtgagttacagatgtagacaaagcaggagatagaacattcaat 2401
Db 2386 CGATCATAAATTTAAGATAGGTACAGATGTAGAGAAAGAAAGAGATGGATACACACAGT 2445
QY 2402 ctgcataatgggaagtggatctgtaacccaaactgtacaag - - - - -caaaagattctcg 2455
Db 2446 CTCATAAATGGAAGATGGATCCAGAGTGAGTCACTCTCAATGGCAACAATACAATATG 2505
QY 2456 cctctctccccagatctccaaatgccaaatgccaaatgtagtgaacacacccgtgaaacttgya 2515
Db 2506 CCACCTCCACCTCAAGATCCCAATTTCTCACATATGACAAACCCACACTGAATTTATCGGA 2565
QY 2516 tggagaaaaagtatctgttcttggcaagatggttaaccttaactcagggcccaagaagaat 2575
Db 2566 TGGAGAAAAGTATCTGTCTTGGCCAGAAAATATCATTTACAGAAAGGAGGAAGAAAT 2625
QY 2576 ggtgtgaacatgggaagtggcagtcgttacacgctgcacggaaaaaatlccatlytic 2635
Db 2626 TACATGCAAAAGATGGAAGATGGCAGTCAATACCACCTCTGTGTGAAAAAATTCATGTTC 2685
QY 2636 ccagccctcaaatgaacatggatcttattgaatgcgccaggtctctcgagagagagag 2695
Db 2686 ACAACCCCTCAGATAGAACAGGAACCATTAATTCATCCAGGTCTTCAAGA - - - - - 2739
QY 2696 agatttaattgagtcagcagttatgaacacggaactacattcagctattgctgtagaga 2755
Db 2740 - - - - -AAGTTATGCACATGGGACTAAATTCAGTTTACTTTGTGAGGG 2781
QY 2756 tggattcaagatatctgaagaaaatagggtgaacctgcaacatgggaaaaatggagctctct 2815
Db 2782 TGGTTTCAGGATATCTGAAGAAAATGAACAAACATGCTACATGGGAAAATGGAGTCTCC 2841
QY 2816 gctcgtgtgtggaataaccttggaccacccttcaattcctctctggtattgttctc 2875
Db 2842 ACCTCAGTGTGAAGGCTTCTCTGTAAATCTCCACCTGAGATTTCTCATGGTGTGTAGC 2901
QY 2876 tcatgaactagaagaattaccaataatggagagaggtttacatacaatgttctgaaggtct 2935
Db 2902 TCACATGTCAGACAGTTATCAGTATGGAGAACAGTTAGTACAAATGTTTTCAGAGTTT 2961
QY 2936 tggaaattgatgacacagcatttataatgtgtlagggagacagtggtctgaaacctcccaa 2995
Db 2962 TGGAAATTGATGGCCCTGCAATTTGCAAAATGCTTAGGAGAAAAATGGTCTCACCCCTCATC 3021
QY 2996 atgcataaaaaactgattgtgacaacttgcaccaatttgaattggcaaacccagacagaaa 3055
Db 3022 ATGCATAAAAAACAGATTGCTCAGTTTACCTAGCTTTGAAAATGCCATACCCATGGGAGA 3081

QY 3056 gaaaaaaaatcaccaggttcagagagaacaagtgacattcagatgtgccacctccgtagc 3115
Db 3082 GAAGAAGGATGTTGTTAAGCGGGTGAGCAAGTGACTTACACTTGTGCAACATATTACAA 3141
QY 3116 aatggatggctctgacattgtcaccatgttataacgaagtggattggacagccggtatg 3175
Db 3142 AATGGATGGAGCCAGTAATTAACATGTCATTAAATAGCAGATGGACGAAGGCCAACATG 3201
QY 3176 caaagataattctcgtgtgaatccaccacatgtgcccataatgctactatacaagaagca 3235
Db 3202 CAGAGACACCTCTCTGTGTGAATCCGCCACAGTACAAAATGCTTTATATATGTGCGAGACA 3261
QY 3236 caagactaaatattccatctggtgacaaagtacgttatgactgtataaaaaccttltgaatt 3295
Db 3262 GATGAGTAAATATCCATCTGGTGAGAGAGTACGTTTATCAATGTAGGAGCCCTTATGAAT 3321
QY 3296 atttggggaagtgaagtgtatgcccacacgggatttggacagaaccccccgaatgcaa 3355
Db 3322 GTTTGGGGATGAAGAAGTGAATGTTTAAATGGAAAATGGACGGAACCCACTCAATGCAA 3381
QY 3356 agattcaacagggaatgtgggctcctccacctattgacaattggagacatcacctcctt 3415
Db 3382 AGATTCTACAGGAATAATGGGCCCTCCACCTATTGACAATGGGACATTTACTTCATT 3441
QY 3416 gtcattaccagtatatgcaccattatcatcagttgaaatcaatgccagaactattatct 3475
Db 3442 CCGCTGTGAGTATATGCTCCAGCTTCATCAGTTGAGTACCAATGCCAGAACTTGTATCA 3501
QY 3476 acttaagggaataaagatagtaacatgtagaatgaaagtggatcagccaccaacctg 3535
Db 3502 ACTTGAGGGTAAACAGCGAATAACATGTAGAAATGGACAATGGTTCAGAACCCACAAATG 3561
QY 3536 ctctacatgcatgtgtatccagaagatatattggaaaaacataatattgttctcagatg 3595
Db 3562 CTATACATCCGTGTGTAATATCCGAGAAATTTATGGAATAATTATAACATACCATTAAGGTG 3621
QY 3596 gagggaaaaatgcaagaattattcccacacagggggagaatattgaattcactgtgtaaac 3655
Db 3622 GACAGCCAAACAGAAAGCTTTATTCGAGAACAGGTGAATCAGTTGAAATTTGTGTAAACG 3681
QY 3656 tggatatagaataatcagaggtcacctcgtttctgacaaaagtcgatlgaggtcacat 3715
Db 3682 GGGATATCGTCTTTCATCAGTCTCACATTTGCCAACACACATGTTGGGATGGGAAACT 3741
QY 3716 caattatccacttgtgtataaaatcgctatacaattattagtaaaccttattgagtaga 3775
Db 3742 GGAGTATCCAACCTGTGCAAAAAGATAGAATCAATCAATAAAGTGCACACCTTTATTTCAGA 3801
QY 3776 aatgcacatgtatatactaataacagtttgaatttcaattttaaatttattgttagctcatl 3835
Db 3802 ACTTTAGTATTAAATCAGTTCTCAATTTTCATTT - - - - -TTTATGTATTGTTTACTCCTT 3856
QY 3836 tctctcaataagata 3851
Db 3857 TTTATTTCATACGTAA 3872

RESULT 6

MUSCFHRD

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

2955 bp mRNA linear ROD 12-JUN-1993
Mouse complement factor H-related protein mRNA, complete cds, clone 9C4.

M29009 J05259

M29009.1 GI:192563

complement factor.

Mouse (strain C57/B10.WR) liver, clone 9C4.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 2955)

Vik,D.P., Munoz-Canoves,P., Kozono,H., Martin,L.G., Tack,B.F. and

Chaplin,D.D.

TITLE Identification and sequence analysis of 4 complement factor
 JOURNAL H-related transcripts in mouse liver
 MEDLINE J. Biol. Chem. 265, 3193-3201 (1990)
 COMMENT 90153969
 Draft entry and computer-readable sequence for [J. Biol. Chem.
 (1990) In press] kindly submitted
 by D.P.Vik, 13-OCT-1989.
 FEATURES
 source Location/Qualifiers
 1..2955
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 69..2495
 /note="complement factor H-related protein"
 /codon_start=1
 /protein_id="AAA37416.1"
 /db_xref="GI:309166"
 /translation="MGFCSMLLSNLIITAWLSTAKGEVKSCEFPQKYGRLYFEBIL
 RNPFPVSIKNKYSKCDNGFSPGLFDWYLRCTVQGWKPEVPCVRKGVFHYVENG
 AYWEKLYVOGSLKVOQYNGYSLONGQDMMCTCTENGWSPPKCIRIKTKCSVDIEIEN
 GFSESFRTYALNRETSYRCKOGYVNTGETSRSTICLONGWSPQSKISCEKRPVFE
 NSVTKNNSWFLNKLDRLECLIGHENYKHTKGSITCTYIGMSDTPSCYEIECSVPI
 LDKLVSPRKEKRYVDLLEFVSCRGHRVSDVQCYHFGWSPSPPTTKGVAVESCAQ
 PPEIPNGEINGKRYFVSHGEVGVGDKPFLKGPNGKIQCVDSGMSPTTKGVAVESCAQ
 CGDIELEHSGVKFSVPYHHGDSVEFTCAETFTMIGLSVSLGSKWTOLPKCVATD
 QLEKRVLSKTDIEAIKPKRNEFOHNSWYKCRKOQEVESHICINGKNWSPNCTRK
 TSCGPPPOIPLTVLETVKYLDEKLSVLCDQNYLTQDPEEMCKDGRWQSLPHCIG
 LFCGPPPSILRTVLSLELSYHGEVTVHCSFGIDGPAFKCEGGKWSDDPPKCIK
 TNCVDLITENAIIRKKKSYRTGEVTVFCOSPYOMNGSDTVTCVNSRWIKGPVKC
 DSRGCKGPPPDINDGIDITSLSPLEPEFSSVDYQCKYLLKGGKTIITCRNGKWEPP
 TCLHACIIPENMEARKIILKRWHTENIYSHGSDIEPECKYGYHKARSGSPPTKCI
 SGTINVPTE"
 BASE COUNT 990 a 550 c 577 g 838 t
 ORIGIN
 Query Match 38.2%; Score 1616; DB 10; Length 2955;
 Best Local Similarity 73.3%; Pred. No. 0;
 Matches 2375; Conservative 0; Mismatches 465; Indels 400; Gaps 8;
 QY 964 ttactgctggtccctgctccaaagttagcttgaaaccttgatatttcccaattca 1023
 Db 109 TAACTGCATGGCTTTCCACTGCTAAAGGGGAAGTGAATCTTTGTGAATTTCCAATAATCA 168
 QY 1024 aacatggacgtctgtattatgaagaagccggagacccactactlcccgactacacagtcatactggact 1143
 Db 169 AATATGACGCTGTATTTGAAGACATCTTGACACCCCACTTCCACGATCTATAGGAA 228
 QY 1084 aggagtacagctataacttgacaacggttttacaacgccttcacagtcatactggact 1143
 Db 229 ATAAGTACAGCTATAAGTGTGACACGGGTTTTCACCACCTTCTGGGGACT 288
 QY 1144 acctctgttcacagtaaatgggtggagccctgaagcttcacgtcctcaggcaatgtattt 1203
 Db 289 ACCTTCGTTGCACAGTACAAAGGTTGAAGCCCTGAAGTCCCATCTGTGCAGGAATGTGTTT 348
 QY 1204 tccattatggaatatgagaattcttactactggcaagaagatatagagggctcagt 1263
 Db 349 TCCATTATGTGAGAATGAGAATTTGCATCTGGGAAAAAATATATGTGACGGGTCACT 408
 QY 1264 ctgcaaaagtcaggtcacaatggctatagcttctcccaaatggctcaagatacatattt 1323
 Db 409 CTTTAAAGCTCCAGTGTATAATGCTATAGTCTTTCAAAATGGTCAAGACACAATGACAT 468
 QY 1324 gtacagagaatggctgggtccctcctcccaaatggctcgatcagcactgttccagat 1383
 Db 469 GTACAGAGAATGGCTGGTCCCTCTCCCAATGATGATCGGTATCAAGACACAATGACAT 528
 QY 1384 cagatatgaattgaataatgggtttttttctgaaatctgattatcacatagctcctaaata 1443
 Db 529 CAGATATAGAAATTTGAGATGGGTTTTTTTCTGAAATCTTTTCGTACATATGCTCTAAATA 588
 QY 1444 gaaaacacggttatagatgtaaacaggatatgttaacaaaataccggagaataatcagaa 1503
 Db 1669 AATTATCTGTTCTTTGCCAAGACAATTAACCTAAGTGTGAAATCTTTGGATGGGAGAA 1668

Db 589 GAGAAACATCTCTATAGATGTAACACAGGATATGTGACAAATACTGGAGAAACGTCAGAT 648
 QY 1504 taattactgtcttcaagatggatgggtcacctcgacccctcatgattaaagtctttgtgata 1563
 Db 649 CAATAAATCTGCTTCAAAATGGATGGTCACCTCAACCTCATGTATTAAAGTCTTTGTGAAA 708
 QY 1564 tgctgtatttgagaattcttactaagaataataacacatggctttaaactcaatgaca 1623
 Db 709 GACCAGTATTTGAGAAATTTCTGAACTAAGAAATAATAGTACATGGTTTAAAGCTCAACGACA 768
 QY 1624 aattagactatgaatgtcacatiggtatgaaatgaataataacacatacacaaggtccta 1683
 Db 769 AATTAGACTATGAATGCTCATTTGGACATGAATAATGAATAATAACATACCAAGGCTCTA 828
 QY 1684 taacatgtactattgatggatggcttagtcacacctcctgttatagaagagaatgcagca 1743
 Lu 829 TAAGGTGCTACTTATTATGGATGCTGCTATACGCCCTCATGTTATGAATAGATGCGAGCG 888
 QY 1744 tccccctgttacaccaagaattagttgtttttccccagagaagtaataacaaagtgtgag 1803
 Db 889 TTCCCATCTTAGACCGAAAACTAGTCGTTTCTCCAGAAAAAATAACAGATTTGGAG 948
 QY 1804 attcgttgagttctctctgctggttcagacacagagttggagcagattagtcacatgct 1863
 Db 949 ATTGTGTAGAAATTTCTCTGCGCTTCAGGACACAGAGTTGGGCCAGATTCAGTGCATGTT 1008
 QY 1864 accactttggtatggctcccttaatttccccacgtgtgaagcccaagtaaaatcatgtgacc 1923
 Db 1009 ACCACTTTGGATGGCTCTCTAGTTTCCCTACATATAAAGGTCAAGTAGCATCATGTGCAC 1068
 QY 1924 aacctctgaaatccccgaattggggaaataaagggaacaaataaagtgtgaatacagccatg 1983
 Db 1069 AACCTCTGAAATTCCTAATGGGGAAATTAATGGAGCAAAAGAGTTGAATACAGCCATG 1128
 QY 1984 gtacagtggtggaatgatgattgcaaacctagatttctactgaaggaacccaataaaatcc 2043
 Db 1129 GTAGGTGGTGGATATGATTTGCAAACTAGATTCTTACTGAAGGGACCCCAATAAATATCC 1188
 QY 2044 agtgtgttgacgggaagtggacaaggttgcgcgatatgcgttgagttgagtagagaacatgtg 2103
 Db 1189 AGTGTGTGTATGGATGCTGGACAACCTTGCTCTGTATGTGTGAGGAGGACAGAACATGTG 1248
 QY 2104 gagaacctcctgaacttgagcagctgctctgcaagttatctgtccctccctacacatcag 2163
 Db 1249 GAGACATTTCTGAACTTGAACATGGCTCTGTCAAGTTTCTGCTCCCTCCCTACACCATG 1308
 QY 2164 gacattcagtggaagttcactgttacagaaccttcaaatgattgacatgacatgagttt 2223
 Db 1309 GAGATTCAGTGGAGTTCACTTGTGCAAGAACCTTCAACATGATTGGACTTGGGTTCAGTTT 1368
 QY 2224 tctgcattagtgggaagtggaacccagcttctcctcaatgtgttgcaacagatcaactggaga 2283
 Db 1369 CTTGCTCTAGTGGAAAGTGGACCCAGCTTCCCTAAATGTGTGTGCAACAGACCAACTGGAGA 1428
 QY 2284 agtgtaaagccccgaatgcaactgagcagatgacatgcaactcacaataagaatgaattta 2343
 Db 1429 AGTGTAGATGCTGTAAGTCACTGACATAGAAGCAATTAACCAAAAGAAAGAAATTC 1488
 QY 2344 atcataacttttagtgagttacagatgtagacaaaagcaggagtagtgaacatcactc 2403
 Db 1489 AGCATAAATCCACCATGTATTACAAATGTAGACACAGCAGGAGTATGAACACTCAATCT 1548
 QY 2404 gatcaatggaagatgggtatcctgaaccacaaactgtacaagcaaaagattcttgcctccctc 2463
 Db 1549 GTATCAATGGAAAAATGGAATCTCTGAACCAAACTGTACAAGGAAAAACATCTCTGCCCTCCTC 1608
 QY 2464 ccccgacagattcccaatgcccgaagtattgaacacccgtgaataacttggatggagaaa 2523
 Db 1609 CACCACAGATTTCCAAATACCTTAGTGTGAAATCTGTGAAATCTTTGGATGGGAGAA 1668
 QY 2524 aggtatctgttcttggcgaagtgttacctaacctcagggccccagaagaatgggtgtga 2583
 Db 1669 AATTATCTGTTCTTTGCCAAGACAATTAACCTAAGTGTGAAATCTTTGGATGGGAGAA 1728

QY	2584	aacatggaaggtggcagtcggttaccacgctgcacggaataattccatgttccccagcccc	2643
DB	1729	AAAGATGGAAGGTGGCAGTCATCTACTCACTGCA	1761
QY	2644	ctaaattgaacatggaatctatttaagtgcgccagctcctcagaagagagagalltaa	2703
DB	1762	-----	1761
QY	2704	ttgagtcagcagttatgaacacggaactacattcagctattgtctagagatggallca	2763
DB	1762	-----	1761
QY	2764	agatatctgaagaaatagggtaacctgcacatgggaaaaatggagctctcctgcctcgtt	2823
DB	1762	-----	1761
QY	2824	gtgttgaataccttgtggacccccacccttcaattcctcttggtattgttctcatgaac	2883
DB	1762	--TTGACATCTCTTGTGGACCTCCACCTTCAATTCTTCTGGTACTCTTCTTCTTGACG	1818
QY	2884	layaaagttaccaatatgagagagaggttaccatacaaatgttctgaagcgttttgaattg	2943
DB	1819	TAGAGATTACCAACATTTGGGGAAAGAGGTTCACATACCATTTGTCTACGGCTTTGGAATTG	1878
QY	2944	atgaccagcatattataaatgtgtagggagacatgctctgaacctccccaaatgcataa	3003
DB	1879	ATGGACCAAGCATTTATTAAATGTGAAGGAGAAAGTGGTCTGACCCACCAAAATGCATAA	1938
QY	3004	aaactgatttgacaactgtccacatttgaattgccaacccgcacagacgaagaaaaaaa	3063
DB	1939	AAACGAATTGTGACGTTTTTACCACCAATTTGAAATGCCATATATAGAGCAAGAAAGAAA	1998
QY	3064	aatacacaggtcagagaaacaagtgcactcagatgtccacctcgtatcgaatgaatg	3123
DB	1999	AATCATATAGCAGCAGGAGAACAAAGTGACATTCAGATGTCAATCTCTTATCAAAATGAATG	2058
QY	3124	gctctgacattgtcacatgtgttaatacgaagtggaatggacagcgcgtatgcgaagata	3183
DB	2059	GCTCAGACACTGTGCATGTGTTTATAGCGGTGGATTTGGAAGGCCAGT-----	2107
QY	3184	attcctgtgtaatccaccacatgtgccaaatgtcactatactatacaagggcacagaacta	3243
DB	2108	-----	2107
QY	3244	aataccatctcgttgacaaagtacgttatgacgtgaaataaacctttgaaattatllgggg	3303
DB	2108	-----	2107
QY	3304	aagtggaaagtgtgtgccaaaacgggatttggacagaaccacgaaatgcaaaatlcac	3363
DB	2108	-----ATGCAAAAGACTCAA	2121
QY	3364	cagggaatgtgggacctctccacctattgacaatgagagacatcacctccttgcattac	3423
DB	2122	GAGGGAAATGTGGGCTCTCCACCTATTTGACAATGGAGACATCACCTCTTGTCAATTAC	2181
QY	3424	cagtatatcaccattatcatcagttgaatatcaatgccagaactattatctactlaagg	3483
DB	2182	CAGAAATGAAACCATTTCTCATCATGTTGACTATCATGTGCCAGAAGTATTATCTCTTTAAGG	2241
QY	3484	gaataagaagtagtaacatgtagaaaatggaaagtgctcagccaccaacctgtctacatg	3543
DB	2242	GAAAGAAGACAATAACATGTAGAAATGGAAAGTGGTCTGACGCCACCACATGCTTTACATG	2301
QY	3544	catgtgtatcaccagaagatatattggaaaaacataaatatagttctcagatggaggaaa	3603
DB	2302	CATGTATAATPACCAAAAAACATTATTTGGAAGCAGCAGCAAAATAATTCTTAATGGAGACACA	2361
QY	2604	atgcaagatttattcccacatcaggggagaataattgaattcatgtgaaaccttggatata	3663
DB	2362	CTGAAATATTTATTTCCCAATTTCAGGGAGGATTTTGAATTTTGAATTTGAATTTGATATC	2421

QY	3664	gaaattcagaggaacacccctcgcttcgtacaaagcgtatgaggggtcacatcaattatc	3723
Db	2422	ATAAACAAGAGGATCACCAGCGCATTCGTACAAAGTCGATAGTGACACCAATCAATATC	2481
QY	3724	ccacttggtataaaatcgctatacaaatattagtaaaaccttatgtagagaaatgcaca	3783
Db	2482	CCACTTGTGAATAAATCGAATCATATTTATTAGTTGATTTTATTGTTTACGAAAGACACA	2541
QY	3784	tgtattactaatacagtttgaatttacattt-aaatattggttagctcatcttcctccta	3842
Db	2542	AGCATGTCATAATACTATCTTCAATTTGCATTTGAAATATTGTTTAACTCATGCTCTCTC	2601
QY	3843	ataagtataaaacttttttatactgggtggaataacagtaaacctttacagagactggtggcac	3902
Db	2602	ATAAATATAAACA-TTTTGTGTATATGGTGATTAATTTTAACTTTAAAGATGATGGCCAA	2660
QY	3903	aaagcaagaacattacattcaaaacctcctaataaccaaa-tatgatattgccaaaggacaaac	3961
Db	2661	ATGCAATGCAGTACATCAAAACTCCTAATCCAAACCATTTATATGTCGAAGGACAAC	2720
QY	3962	tatgtctaaagcaaaaaataatgtagttcttctaatgctgtcttttttttttcaggagacctt	4021
Db	2721	T-----CAAATAAAGAAAAATTCATGATGTCTTTTCAGCATGTTGTTTCTATTACAGACCTCT	2775
QY	4022	cagatttcttgataacctttt---tgtaggtctgattccacagtgagtggaagacacac	4078
Db	2776	TCAGATTTCCTAGATATCTCTTTGATGTAAATGTTTGTATTTATAGTGAATGAAGAATATAT	2835
QY	4079	tgactctgactcacaattagtaattacttgc--aatcacattaaacacaaacatcatcaat	4136
Db	2836	TGACTCATCTTCCAATTAATATGATTTCCCAAGCATGTAAAGAACCAACTATCATATA	2895
QY	4137	atcacaaatgtatacagctaaattactgtgtctacctttgtatcatcaataaagaatctaag	4196
Db	2896	TTATATCACTAATGCATATAATTAATCACTATATAACTTTTCAAAATAAAGAATCTAAG	2955
RESULT	7		
AF436847		2729 bp	linear
LOCUS		Rattus norvegicus complement factor H-related protein mRNA,	31-OCT-2001
DEFINITION		complete cds.	
ACCESSION	AF436847	GI:16551387	
VERSION	AF436847.1		
KEYWORDS		Norway rat.	
SOURCE		Rattus norvegicus	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
REFERENCE		1 (bases 1 to 2729)	
AUTHORS		Ren, G. and Quigg, R.J.	
TITLE		Rat complement factor H-related protein sequence	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 2729)	
AUTHORS		Ren, G. and Quigg, R.J.	
TITLE		Direct Submission	
JOURNAL		Submitted (19-OCT-2001) Nephrology, University of Chicago, 5841 S. Maryland Ave., Chicago, IL 60637, USA	
FEATURES		Location/Qualifiers	
source		1..2729	
		/organism="Rattus norvegicus"	
		/strain="Sprague-Dawley"	
		/db_xref="taxon:10116"	
		/cell_type="glomerular epithelial cells"	
		/tissue_type="kidney; also found in liver"	
		50..2149	
		/codon_start=1	
		/product="complement factor H-related protein"	
		/protein_id="AA25802.1"	
		/db_xref="GI:16551388"	
		/translation="MRFYSIWLIIALLISLTAKGEECVLIHYVNGYSLFLERK"	
		VQDSRVKVCYKGFNLPNGDITICTESGTPPICIPSEKSKSDIEIDNGFFSRSI	
CDS			

LTVAVNRKTRYCKQGYTANGERSGSLTCLQTGWSQAQPSICSDVPVPEFENAKSND
STWFKLNDVDECHGYENKYHTKTKGSLMCTDQWSKDSQYIECSITLIDPLRV
YPRKITYVDGLDKSCFRGHRVGSQVQYDFGNSFPTCKQGSQYQYPELLNG
IEGTGKFAFVYHGDVVEYDCKPRLFLKNGICQVCKQWYLTPTCEBKTKQDPEL
EHGSVRFPVPHSHGVSVEYTCAEFTWIGOGSVYCSIRGWAOLPCRVATDLEKKA
PLRAVLTGKLLKGLCNHNSLSYKCRSGQKXGYSCINGRWDPEPTCTKREKTCFPP
PQOIPNANHETTVRYLQEKVSLVQCYLQGEFENWKGHRWSLPRCTKWS
QPKRDIHGSILKPLRSDGDAQRYEYSHHGHTTYSVCDGFRKSENGVYCHGMS
SPRQVCEKGCQPTTEINGDISLSPYAPLUSVEYQCYFVYKMGSKTKCTKNG
EWPEPKCHLACVLTENMRQNIIPKTKNKKYIOSGDFVVEFYVYAKOAGSP
LTVAVNRKTRYCKQGYTANGERSGSLTCLQTGWSQAQPSICSDVPVPEFENAKSND

BASE COUNT	916 a	502 c	561 g	750 t
ORIGIN				

Query Match 29.1%; Score 1229.4; DB 10; Length 2729;
Best Local Similarity 83.6%; Pred. No. 8..le-264;

	Matches	1408:	Conservative	0:	Mismatches	271:	Indels	6:	Gaps	1:
Qy	1193	gcaatgtatttccattatgttgaatatggaagaatcttctatactggaagaagaatatat	1252							
Db	124	CGAATGTCTTTTGTGCAATATGTGGAAATGGTTATCTCTTATTTCTCTGGAAGAAATTTGT	183							
Qy	1253	agagggtcagtcgcaaaagtcagtgctcagtggtctatagcttccaaatggtcaaga	1312							
Db	184	ACAAGATCAGTCTGTCAAGTCACTGTTTAAAGGCTTCAACCTTCCAAATGGTCAGGA	243							
Qy	1313	tacatattatgttacagagaatgggtggctccctctcccaaatcgctccgtatcaaacg	1372							
Db	244	TACAATTATATGTACAGAGAGTGGCTGGACACCTCAACCCATATGCATCCCTCTTGAGAC	303							
Qy	1373	ttgttcagtatcagatatagaaatgaaaaatgggttttttttctgaatctgattatacata	1432							
Db	304	ATGCTCAAGTCAGATATAGAAATGTAAATGGATTTTTTTCTGAACTGTGATTTAAACATA	363							
Qy	1433	tgctctaaatgaaaaaacacgggtatagatgtaaacagggtatgttaacaaataccggaga	1492							
Db	364	TGCTGTAAATAGAAAAACACGATATGTGTGTAACAGCGGATATACACAGCAAAATGGAGA	423							
Qy	1493	aatcacggaataattacttgtcttcaagatggatgggtccactgcagccctcatgattaa	1552							
Db	424	AAGATCGGGATCAATCACTTGCCTTACAACHTGGATGGTCTGCTCAACCTTCTTGCAATTAA	483							
Qy	1553	gtcttgtatgatgctgttatttgagaattcttatgactaagaataataacacatggtttaa	1612							
Db	484	GTCVTGTGATGTGCTGTGTGTGAAATGCTTAAATCTAAAAATGATACGACATGGTTTAA	543							
Qy	1613	actcaatgacaaattagactaatgactgcacatggatatagaataaataaacataac	1672							
Db	544	ACTCAATGACAAATTAGATTAATGATGTCACTTGGATATGAAATAAATAATAAGCATAC	603							
Qy	1673	caaaggctctataacatgtacttatgatggatggcttagtcagcacccctctgttatgaag	1732							
Db	604	CAAAAGGCTCCATAATATGTGTACTACTGTATGGATGGTCTGTATAAACCTCATGTTATGAAT	663							
Qy	1733	agaatgcagcaattccocgtttacaccaagactttagttgtttttccagagaagataaata	1792							
Db	664	TGAATGCAGCATTTACCATATTAGACCCACGCTTAGTTGTTTATCCCAGGAAATATAATA	723							
Qy	1793	caaagttggagattcgttgagtttctcttcgcttcaggacacagagtgtggagcagattt	1852							
Db	724	TACAGTTGGAGATTGTTGAAATTTCTCTTGCCGACCAGGACACAGAGTGGACCAGATTTC	783							
Qy	1853	agtcaatgctaccacttttgatggtgctcccttaatttcccaacggtgaagcccaagtaaa	1912							
Db	784	AGTCCAATGCTATGACTTTGGATGGTCCCCTAGCTTTCCGACATGTAAAGGTCGAAGTAAG	843							
Qy	1913	atcatgtgacaaacctcttgaaatccccgaatgggggaataaagggaacaaaaaagttga	1972							
Db	844	ATCATGTGGTCAATATCCTGAACCTACTCAATGGGGAAATTAAGGGACAACAAGAAGCTGA	903							
Qy	1973	atacagccatgggtgaactggttgaatatgattgcaaaccttagatttctactgaaggacc	2032							

RESULT	8
A91886	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

A91886 1532 bp DNA linear PAT 22-JAN-2000
Sequence 4 from Patent WO9823638.

VERSION	A91886.1	GI:6740763
KEYWORDS		
SOURCE		unidentified.
ORGANISM		unidentified
		unclassified.

```
REFERENCE 1 (bases 1 to 1532)
AUTHORS Schwaebler,W. and Sim,R.B.
TITLE COMPLEMENT INHIBITOR
JOURNAL Patent: WO 9823638-A 4 04-JUN-1998;
          SCHWABLER WILHELM (GB); UNIV LEICESTER (GB)
FEATURES             Location/Qualifiers
     source          1..1532
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT   523 a   275 c   290 g   444 t
ORIGIN
Query Match      26.5%; Score 1119; DB 6; Length 1532;
Best Local Similarity 100.0%; Pred. No. 3.6e-239;
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3111 tatcgaaatgagctgctgacattgtcacatgtgtaataacgaagtggattgacagccg 3170
DB 414 TATCGAATGGATGGCTCTGACATGTGCACATGTCTTAATACGAAGTGGATTGGACAGCCG 473

QY 3171 gtatgaaagaataatctctgtgtaataccaccacatgtgccaaatgctactataca 3230
DB 474 GTATGCAAAATTAATCTCTGTGTAATCCACCAATGTGCCAATGCTACTATACTAACCA 533

QY 3231 aggcacaagactaaataccatctggtgacaaagtcattatgactgtaalaacactttt 3290
DB 534 AGGCACAAGACTAAATATCCATCTGTGTGACAAAGTACGTTATGACTGTATAAACCPTTT 593

QY 3291 gaattatttgggaagtgggaagtgtgacaaacgggatttggacagaaacacagaaa 3350
DB 594 GAATTATTTGGGAAGTGGGAAGTCATGTGCCAAACGGGATTTGGACAGAACCCAGAAA 653

QY 3351 tgcaaaagattcaacagggaaatgtggccctctccactattgacaatggagacatcacc 3410
DB 654 TGCAAAAGATTCAACAGGAAATGTGGCCCTCTCCACCTATTGACATGAGACATCACC 713

QY 3411 tccttgcataccagtatatgcaccattatcatcagttgaaatcaatgccagaactat 3470
DB 714 TCCTTGTTCATTACAGTATATGCACCATTTATCATCATGTTCAATATCAATGCCAGAACTAT 773

QY 3471 talctacttaaggaaataagatagtaaacatgtagaataagaaagtggtctcagccacca 3530
DB 774 TATCTACTTAAGGAAATAGACATAGTAACATGTAGAAATGGAAGTGGTCTCACCACCA 833

QY 3531 acctgttacatgcatgtgtgataccagaagatattatggaataataatagttctc 3590
DB 834 ACCTGCTTACATGCATGTGTATACCAAGAGATATTATGCAAAACATAATATAGTTCTC 893

QY 3591 agatggaggaaataacaaagatttattcccaatcaggggagagaatatgaaatcatglt 3650
DB 894 AGATGGAGGAAATGCAAGATTTATTCCTCAATCAGGGGAGAAATPTGAAATTCATGTGT 953

QY 3651 aaacctggatagaaaaattcagaggatcacctccgtttcgtacaaagtgcattgaggt 3710
DB 954 AAACCTGGATATAGAAATTCACAGGATCACCTCCGTTTGTGACAAAGTGCATTGAGGT 1013

QY 3711 caatacaattatccactgtgtataaaatcgctatacaattattagtaaaccttatgga 3770
DB 1014 CACATCAATTTATCCCACTTGTGTATATAAAATCGCTATACAAATTTATAGTAAACCTTATGGA 1073

QY 3771 tggaaatgcacatgatatattactaataacagtttgaaattacatttaaatattgttagc 3830
DB 1074 TGAAATGCACATGTATATTACTAATACAGTTTGAAATTTACATTTAAATATTGTTTACG 1133

QY 3831 tcatttctctaataagatatataaaacttttttatatggtgggttaacagaaacttaca 3890
DB 1134 TCATTTCTCTAATAAGTATATAAACTTTTATATGTTGGTTTATCAGTAACCTTATACA 1193

QY 3891 gactgttgcacaaagcagaacattacattcaaaactcctaataccaaatgatgttc 3950
DB 1194 GACTGTTGCCAACAAAGCAAGAACATTACATTCAAACTCCTAATCCAAATATGATATGTC 1253

QY 3951 caaggacaaactatgtctaaagcaagaaaaataatgttagttcttcaatgtctgttttat 4010
DB 1254 CAAGGACAAACTATGCTAAGCAAGAAAAATAATGTTAGTTCTTCAATGCTGTTTTAT 1313

QY 4011 tcaggaccttcagattttcttggatacccttttggtaggttctgattccacagtgagtgga 4070
DB 1314 TCAGGACCTTTCAGATTTTCTTGGATACCTTTTGTAGTTCTGATTTCACAGTGAGTGGA 1373

QY 4071 agacacactgactgacttcaaatagattattactgcaatacattacaacaaactat 4130
DB 1374 AGACACACTGACTCTGACTTCAAAATTAGTATTACTTCCAATACATTAAACCAACCAACTAT 1433

QY 4131 cataatatcacaatgtatcacagctaatattactgtgtcctacccttggatcaataaagaaa 4190
DB 1434 CATAATATCAAAATGTATACACTAATTAATCTGTGCTCTACCTTTGTTATCAATAAGCAA 1493

QY 4191 tctaaagaagtcttctgcttaaaaaaataaaaaa 4229
DB 1494 TCTAAGAAAGTTCTTGTCTTAAAAAATAAAAAA 1532

RESULT 9
HUMCHF HUMCHF 2132 bp mRNA linear PRI 01-NOV-1994
LOCUS Human complement H factor mRNA, complete cds.
DEFINITION M17517
ACCESSION M17517.1 GI:180497
VERSION H factor; complement; glycoprotein.
KEYWORDS Human liver, cDNA to mRNA, clone R2a.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2132)
AUTHORS Day A.J., Ripoch,J., Lyons,A., McIntosh,B., Harris,T.J. and
          Sim,R.B.
TITLE Sequence analysis of a cDNA clone encoding the C-terminal end of
          human complement factor H
JOURNAL Biosci. Rep. 7 (3), 201-207 (1987)
MEDLINE 88025472
FEATURES             Location/Qualifiers
     source          1..2132
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="1q32"
     gene            1..2132
                     /gene="HF"
                     <1..1976
     CDS              /gene="HF"
                     /note="complement H factor"
                     /codon_start=2
                     /protein_id="AAA52016.1"
                     /db_xref="GI:180498"
                     /db_xref="GDB:G00-120-041"
                     /translation="DFRNVDPKDKQYKVEVLKFSCKPGFTTIGVPSNVQCVHFLGSP
                     DLPCKEQVQSCGPPPELLNGVKNKTEKEEYGHSEVEYICNPRFLMKGNKIQCVDG
                     EWTTLPGVIEESTGCDIPLEHGWALQSPYYGDSVFNCSESTFIGHRSITCI
                     HGVWTLQPCVAIDKILKKSSNLIILEHLKKKFDHNSNI RYRCRKGWIHTVC
                     INGRWDPEVNCMAQIQLPQPPPIPNSHNMTTLN YRQGEKVSVLQCNLYIQEGEE
                     ITCKDGMQSIPLCYVEKIPCSOPQIIEHGTINSRSSQESYAGHTKLSVTCGGFRIS
                     ENEFTTCYMKWSSPQCEGLPCKSPPEISHGVWAHMSDYQYGEVTKCEPGEIGD
                     GPAIAKCLGEKWSHPSCIKTCLSLPSENAIPMEKKADVTYKAGEQVITTCATYTKM
                     DGAINVTCINSRWTGRPTCRDTSVNPPTVQNAVIVSRQMSKYSGEVRVYQCRPYE
                     MFGDEVMLNGMWTPEPQCKDSTGKCPGPPIDNGDITSFPLSVAPASSVEYQCON
                     LYQLGKNRITCRNGQWSEPPKCLHPCVISREIMENYNIALRWTAQKQLSRTGESVE
                     FVCKRGYRLSSRSHTLRTTCWDGKLEYPTCAKR"
BASE COUNT   714 a   385 c   454 g   579 t
ORIGIN 92 bp upstream of Avail site; chromosome 1q32.

Query Match      24.0%; Score 1016.4; DB 9; Length 2132;
Best Local Similarity 69.5%; Pred. No. 2.7e-216;
Matches 1455; Conservative 0; Mismatches 601; Indels 38; Gaps 4;
```


VERSION					
X98697.1	GI:1419423				
KEYWORDS					
complement factor H.					
SOURCE					
ORGANISM					
Bos taurus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;					
Bovidae; Bovinae; Bos.					
REFERENCE					
1 (bases 1 to 2008)					
Soames,C.J., Day,A.J. and Sim,R.B.					
Prediction from sequence comparisons of residues of factor H					
involved in the interaction with complement component C3b					
Biochem. J. 315 (Pt 2), 523-531 (1996)					
JOURNAL MEDLINE					
96202005					
REFERENCE					
2 (bases 1 to 2008)					
Sim,R.B.					
Direct Submission					
Submitted (21-JUN-1996) R.B. Sim, University of Oxford, MRC					
Immunochemistry Unit, Department of Biochemistry, South Parks Road,					
Oxford, OX1 3QU, UK					
FEATURES					
source					
1..2008	Location/Qualifiers				
/organism="Bos taurus"					
/db_xref="taxon:9913"					
/tissue_type="liver"					
/dev_stage="adult male"					
<1..>2008					
CDS					
/codon_start=1					
/product="CCP modules 3-12, with parts of CCP 2 and 13"					
/protein_id="CAAG67257.1"					
/db_xref="GI:1419424"					
/db_xref="SPTREMBL:Q28085"					
/translation="VSGSPHLAEGNFGEYGAKVYYTCDGYQMVGEMNFRCDTGNTG NDPICEVKVLCPVPENPKIFSDALEPDQEYTYGOVFEKNSVMLDGPDKOIHCS TKGWAAETPKCIEFKDPVILNGQAVLPKATYKONRVOYCAAGFEGYGGRTDVC TGSWAPPTICIEITDDPPRIINGVYPPELSKYRGODKITYECKKGDFEPIRGTDATC TRDWVPVRPCAKPCYPVIKHRLYSYRGVEPARVNQOFVYSKDHHFVPPRSORSW DLHACBGHSPEECCLROCFINYLEGNHOHEEKYLQGETVVRHCYEGYGLNQDNAD TWTCESGWSPPPRCIIRVKTCSNRIRIENGFLSEFTFYTLNKRQTECKPGVTATD GKTSGLITLNQNSAQPVCIKSDRPVFEPKARVKSOGDTWFLNRDLRDLYECVDGYENR DGRTGISYCGQGWDSDAKARECSIPEDMDYLNAYPRKETRYKHGDVLKFCSCQR IMVGAOSVQCIFYFGSPKLPTCKRVKKVSKALPELPNGRKKEIKHKEYAHNVEVYA CNPKFLKSHKTKTCQCVDEERTCEISOLDHGDKVPSVPPIIHGDVSVEFA SCREATFMIGPRITICSGEWTOPPCIATDELKCKGSLFFLL"					
BASE COUNT	629 a	382 c	465 g	532 t	
ORIGIN					
Query Match 23.6%; Score 998.4; DB 4; Length 2008;					
Best Local Similarity 70.3%; Pred. No. 2.8e-212;					
Matches 1399; Conservative 0; Mismatches 576; Indels 15; Gaps 4;					
QY	309	tttgggtcccttagctgagcattggaatgcgaatttgaaattggcgcaaatggttat	368		
Dd	4	TCTGGTTCTCCTCATCTTCGGGAAGAAATCAGTTTCAAATGCTGCAGAAAGTTGTTTT	63		
QY	369	acatgtgatgaagggtaccactattaggatgaaattgattaccgtgatgtgatgcagat	428		
Dd	64	ACGTGTGATCAGGGGTATCAGATGGTGGGTGAGATGAATTCCGTGAATGTGACACAAT	123		
QY	429	gggtggaccaatatgatattccaatatgtgaagtgtggaagtgttcgccagtgcagaactg	488		
Dd	124	GGATGGACCATAATATTCCTCATGTGAAGTTGTTTAAGTCTTTACCAAGTGACAGAACA	183		
QY	489	gagaatggaagaattgtgagtggtgcagccgaaccagacagagaataattatttggacag	548		
Dd	184	GAGATGGAAAAATTTTCATGTGATGCCCTGGACCAACCAAGAAATACCTATGGACAA	243		
QY	549	gtggtacgctttgaatgcgaactccgcttcaagattgaagcagagaagaataactgcac	608		
Dd	244	GTGGTACAGTTTGAGTGTAAATTCAGGCTACATGCTGGATGGACCCCAACAAATACATGC	303		
QY	609	tcatataatggcctctggagcaaatgaaaagccacagttgtgtgaaaaattcttctgctgc	668		

Qy	545	acagtggtgacgcttgaatgcaactccggtctcaagattgaagacagagaataatgca	604
Db	586	ACAACCACTACGGTTGTATGTAACCTACAGCTACAAGATTGAAGGAGATGAAGAATCCA	645
Qy	605	ctgctcaataaaggctctgagagcaatgaaagccacagtgctggaataattcttgcct	664
Db	646	TGTGTTACAGACATGGTGTGTTGGAGTAAAGAGAAACCAAGTGTGTGGAATTTTCAGAA	705
Qy	665	gcacacagcagagtgaaatggagatgggtatatactatgcaacccagtttacaaggagaalga	724
Db	706	ATCCCCACATGTTATAATGGATCTCCTATATCTCAGAAAGATTATTTATAGGAGAATGA	765
Qy	725	aagattccaataataatgaagcaagggttttbtgtcacaaagaaaggggagctgtgctgtg	784
Db	766	ACGATTTCATATAAATGAACATGGGTATGAATACAGTGAAGAGAGAGATGCTGTATG	825
Qy	785	cacgggttctgagtgaaatcctcagcttctctgtgaaagaaatgacatttgaactccaata	844
Db	826	CAGTGAATCTGGATGGCTCGCTGGCTTCATGTGAAGAAATAATCATGTGATAATCCCTTA	885
Qy	845	tattccaatggatctacacacctcacagagattaaacacacagaaatgatgataatcag	904
Db	886	TATTCCAAATGGTGACTACTACCTTTAAGGATTAAACACAGAACTGGAGATGAATCAC	945
Qy	905	atatgaatgtaaaatggcttctatcctcgcaaccgcatcacctgtttcaagtgtaacaat	964
Db	946	GTACCAGTGTAGAAATGGTGTATTCCTGCAACCGGGGAAATACAGCAAAATGCACAAG	1005
Qy	965	tactggctggatccctgtctccaagatgagcttgaaacctgtgatttccacacaatcca	1024
Db	1006	TACTGGCTGGATACCTGCTCCGAGATGTACCTTGAACCTTGTGATTATCAGACATTAA	1065
Qy	1025	acatggacgtctgtattatgaagaagccgagaccctactcccgatcacctatagga	1084
Db	1066	ACATGGAGTCTATATCATCATGAGAATATGCGTAGACCATACTTCCAGTAGCTAGGAAA	1125
Qy	1085	ggagtcacagcttaactgtgacaacggttttacacgccttcacagtcatactgggacta	1144
Db	1126	ATATTACTCCTTACTGTGTGATGACATTTTGAGACTCCGTCAGGAAGTTACTGGGATCA	1185
Qy	1145	ccctgtgacagtaaatgggtggagccctggaagttccctcagtcacgcaaatgtatttt	1204
Db	1186	CATTTCATGTCACACAAGATGGATGGTGGCCAGCAGTACCATGCCTCAGCAAAATGTTATT	1245
Qy	1205	scattatggaatagagaatcttcaactactgcaagaagagatatagaggggtcagtc	1264
Db	1246	TCCTTATTGGAAAATGGATATAATCAAAATATATGGAAGAAAGTTTGTACAGGGTAAATC	1305
Qy	1265	tgcaaaagtcagtgacagtggtgtatagttcttccaaatggtccaagatcacattattg	1324
Db	1306	TATAGAGTTCCTGCTCCATCTGCGTACGCTCTTCCCAAAACGGCAGACCACAGTTACATG	1365
Qy	1325	tacagagaatggctgggttccctcctctcccaaatgcgtccgtatca	1368
Db	1366	TATGAGAATGGCTGGTCTCCTACTCCAGATGCATCCGTCGA	1409
RESULT 12			
HSH36			
LOCUS			
DEFINITION Human mRNA for complement factor H 38-kDa N-terminal fragment.			
ACCESSION X04697			
VERSION X04697.1 GI:31991			
KEYWORDS complement factor H.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (bases 1 to 1427)			
Schulz,T.F., Schwable,W., Stanley,K.K., Weiss,E. and Dierich,M.P.			
Human complement factor H: Isolation of cDNA clones and partial			
cDNA sequence of the 38-kDa tryptic fragment containing the binding			
site for C3b			

JOURNAL	Eur. J. Immunol. 16 (11), 1351-1355 (1986)
MEDLINE	87054207
COMMENT	Clone H-19 codes for the first 108 AA of the 142-kDa fragment in addition to the 289 AA derived from the 38-kDa fragment. An additional A residue at pos. 930 not observed in a recently published cDNA clone caused a shift in reading frame. It represents not a sequencing artefact but is probably due to an error that occurred during reverse transcription. Data kindly reviewed (03-AUG-1987) by Schulz T.F.
FEATURES	Location/Qualifiers
source	1..1427
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
misc_feature	1..867
	/note="38-kDa C-terminal tryptic H fragment"
CDS	join(<1..929,931..1195)
	/note="ORF"
	/codon_start=1
	/protein_id="CAB41739.1"
	/db_xref="GI:4725976"
misc_feature	868..1192
	/note="142-kDa N-terminal fragment"
misc_feature	930
	/note="additional A residue, causing a shift in reading frame"
polyA_signal	1407..1412
POLYA_SITE	1427
BASE COUNT	435 a 254 c 333 g 405 t
ORIGIN	
Query Match 16.4%; Score 693.4; DB 9; Length 1427;	
Best Local Similarity 74.8%; Pred. No. 3.2e-144;	
Matches 882; Conservative 0; Mismatches 296; Indels 1; Gaps 1;	
QY	191 ccttgataccggaacacttggtactattgtaaaagtatgcaagaalggagaatgggtacc 250
DB	3 CTGGGATATAGATCTCTTGGAAATGTAATAATGTTATCGAGGAGGAGAAATGGTTGC 62
QY	251 ttctaaccatcaagagatattcggaagaagccatgtgggcatcccgagagacacacctt 310
DB	63 TCTTAATCCATTAAGGAAATCTCAGAAAGGCCCTGTGGACATCCTGGAGATACCTCTTT 122
QY	311 tgggtccctttagctggcagttggatctggaattggaatttggtgcaagggtatttatac 370
DB	123 TGGTACTTTTACCCTTACAGGAGGAAATGTTTGAATATGGTTGTAAGCTGTGTATAC 182
QY	371 atgtgatgaagggtaccactattagtgaaatgtattaccgtgaatgtgatgcagatgg 430
DB	183 ATGTAATGAGGGGTATCAATTCCTAGGTGAGATTAAATACCGTGAATGTGACACAGATGG 242
QY	431 gtgaccaatgatattccaatatgtgaagtgtgaagtcgttcgagtcagacagaaatgga 490
DB	243 ATGGACCAATGATATTCCTATATGTGAAGTTGTGAAGTGTGTACACAGTACACACCA 302
QY	491 gaatggaagaattgtgagtggtgcagccgaaccagacaggaataatttattggacagt 550
DB	303 GAATGGAATAATTTGTCAGTAGTGCATGAAGAACACAGATCGGGAATACCATTTTGGACAAGC 362
QY	551 ggtacgttttgatgcaactccggtcttcaagattgaaggacagaaagaatgcactgctc 610
DB	363 AGTACGGTTTGTATGTAACCTCAGGCTACAAGATGAAGGAGATGAAGAAATGCATTTGTTTC 422
QY	611 ataaaatggcctctcggagcaaatgaaagccacagtcgtggaataatttctgctgcacc 670

Dh 423 ACACGATGCTTTTGGAGTAAAGAGAAACCAAGAGTGTGGAAATTTTCATGCAAAATCCCC 482
Qy 671 acgagttgaaatggagatggtatatctctgaaacacagttttacaagagaatgaagatt 730
Dh 483 AGATGTTATAATGGATCTCTCTATATCTCAGAAAGATTATTTATAAGCAAAATCAACGATT 542
Qy 731 ccaatataaatgtaagcaaggtttgtgtacaaagaagaagagggatgctgtgtgcacgag 790
Dh 543 TCAATATAAATGTAACATGGGTTATGAATACACATGTAAGAGAGGAGATGCTGTATGCACTGA 602
Qy 791 tctgagtggaatcctcagcctctctgtgaaagaatgacatgtttgactccatatttcc 850
Dh 603 ATCTGGATGGCTCGCTTGGCTTCATGTGCAAGAAATATCATGTGATAATCTTTATATTC 662
Qy 851 aaatggtatctacacacccctcacagattataacacacagaattgtagatgaatcagatga 910
Dh 663 AAATGCTGACTACTCACCTTTAAGGATTAAACACACAGAACTGGAGATGAAATCACGTACCA 722
Qy 911 atgtaaaaatgctctctatctgcaacccgatacctctgttcaaaagtgtaaatattctg 970
Dh 723 GTGTAGAAATGGTTTATCTCTGCAACCCGGGAAATACAGCCAAATGACAAAGTACTGG 782
Qy 971 clggatccctgctcaagatgtagcttgaaacctgtgtgattttccacaaatccaacatg 1030
Dh 783 CTGGATACCTGCTCGAGATGATACCTTGAAACCTTGTGATTATCCAGACATTAACATGG 842
Qy 1031 acgtctgtattatgaagaagcggagacccctacttcccagtcacctataggaagagatga 1090
Dh 843 AGGTCTATATCATGAGAAATATCGTAGACCATACTTTTCCAGTAGCTGTAGGAAATATTA 902
Qy 1091 cagctataactgtga- caacgggtttacaaacccctcacagtcatactgggactaccttc 1149
Dh 903 CTCCATTACTGTGATGAACATTTTGAAGACTCCCGTACAGGAATTTACTGGGATCACATTC 962
Qy 1150 gttgcacagtaaatgggtggagcctgaagttccatgctcagcgaatgtattttccatt 1209
Dh 963 ATTGCACACAGATGATGCTGCCAGCAGTACCATGCTCTCAGAAATGTTATTTCTCT 1022
Qy 1210 atgtggaataggaatcttcatactgcaagaagatatataagaggtcagtcgtgca 1269
Dh 1023 ATTTGGAAATCGATATAATCAAAATTAATGGAAGAAAGTTTGTACAGGGTAAATCTATAG 1082
Qy 1270 aagtcaggtcacagtgctatagcttcccaaatggttcaagatacatattattgtacag 1329
Dh 1083 AGTTCCTGCCACCTGCTAGCTCTCTCCAAAGCCGACAGACAGTTACATGTATGG 1142
Qy 1330 agaagtggtgctccctctcccaaatgctcgtcgtatca 1368
Dh 1143 AGAATGGCTGCTCTCTACTCCAGATGCATCCGTGTCA 1181

RESULT 13

LOCUS A91884 866 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 2 from Patent WO9823638.

ACCESSION A91884

VERSION A91884.1 GI:6740761

KEYWORDS

SOURCE unidentified.

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 866)

AUTHORS Schwaible,W. and Sim,R.B.

TITLE COMPLEMENT INHIBITOR

JOURNAL Patent: WO 9823638-A 2 04-JUN-1998;

SWABBLE WILHELM (GB); UNIV LEICESTER (GB)

FEATURES

Location/Qualifiers

1..866

/organism="unidentified"

/db_xref="taxon:32644"

BASE COUNT 273 a 154 c 204 g 235 t

ORIGIN

Query Match 16.3%; Score 691; DB 6; Length 866;
Best Local Similarity 100.0%; Pred. No. 1.le-143;
Matches 691; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tccagtcacactgctccagatagatccaaagacatgagatgtcagcaagaattatttggc 60
Db 1 TCGAGTCAACTGCTCCAGATAGATCCAAGACATGAGACTGTCCAGCAAGAAATTTATTGGC 60
Qy 61 ttaatatattgagactgtttgttagcagaagattgtaaaagttcctcctcaagagaaatt 120
Db 61 TTATATATTGAGACTGTTTGTGTAGCAGAAGATTGTAAAGGTCTCTCTCCAAGAAATTT 120
Qy 121 cagaaattctctcaggttcggtgtctgaacaactatattcagaagcactcagagcaacct 180
Db 121 CAGAAATTTCTTCAGGTTTCGTGCTCTGAACAACATATATTTCAGAAAGCCTCAGCAACCT 180
Qy 181 acaaatgcgcgcctggatcccaaacacttgggtactattgtataaagtatgcaagaattggag 240
Db 181 ACAAAATGCCGCTGGATACCGAACACTTTGGTACTATTGTAAAGGTATGCAAGAAATGGAG 240
Qy 241 aatgggtactcttaaccatcaagatattgcgaaaagcccatgtggcattcccccag 300
Db 241 AATGGGTACTTCTTAACCATCAAGGATATGTGCGAAAGGGCCATGTGGGCATCCCGGAG 300
Qy 301 acacacccctttgggtcctttagctgagctgtggtggtgaattgttggtaaaag 360
Db 301 ACACACCCCTTTGGGTCTTTAGCTGCGCAGTTGGATCTGAATTTGAATTTGGTGCAAGG 360
Qy 361 tigtattacatgtgatgaagggtaccactatttagtgtaaatgtaccgtgaatgtg 420
Db 361 TTGTTTATACATGTGATGAAGGGTACCAACTATTAGTGAATTTGATTTACCTGTGAATGTG 420
Qy 421 atgcagatgggtggaccatgatattcccaatgtgaaagtgtgaaagtgtgtccagtgga 480
Db 421 ATGCAGATGGGTGGACCAATGATATTCCAATATGTCAAGTTGTGAAGTGTCTGCCAGTGA 480
Qy 481 cagaactggagaatggaagaattgtgagtggtgagcccaaccagaccaggaattattt 540
Db 481 CAGAAGTGGAGAATGGAAGAATTTGAGTGGTGCAGCCGCAACACAGACAGGAATATTATT 540
Qy 541 ttggacaggtgtacgctttgaaatgcacactcgggttcaagattgagacagagaagaaa 600
Db 541 TTGGACAGGTGTGTACGCTTTGAAATGCAACTCCGCTTCAAGATTGAAGGACAGAGAA 600
Qy 601 tgcactgctcataaaatggcctctgagcaatgaaagccacagtggtgtggaatttctt 660
Db 601 TGCATGCTCATAAATGGCCTCTCGAGCAATGAAGGACAGACAGTGTGTGGAATTTCTT 660
Qy 661 gctgcccaccacaggttgaaatggagatgg 691
Db 661 GCCTGCCACACAGAGTTGAAATGGAGATGG 691

RESULT 14

MUSCFHRC

LOCUS

DEFINITION

Mouse complement factor H-related protein mRNA, complete cds, clone 3A4/5G4.

ACCESSION M29010 J05259

VERSION M29010.1 GI:192561

KEYWORDS complement factor.

SOURCE Mouse (strain C57/B10.WR) liver, clone 3A4/5G4.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 2356)

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

JOURNAL Unpublished (1989)

REFERENCE 2 (bases 1 to 2356); 1 to 875; 1028 to 2356)

AUTHORS Vik,D.P., Munoz-Canoves,P., Kozono,H., Martin,L.G., Tack,B.F. and Chaplin,D.D.

TITLE Identification and sequence analysis of 4 complement factor
H-related transcripts in mouse liver
J. Biol. Chem. 265, 3193-3201 (1990)
MEDLINE 90153969
COMMENT Draft entry and computer-readable sequence for [J. Biol. Chem.
(1990) In press] kindly submitted
by D.P.Vik, 13-OCT-1989.
FEATURES
Source Location/Qualifiers
1..2356
/organism="Mus musculus"
/db_xref="taxon:10090"
-74..875
/note="Complement factor H-related protein"
Join(74..875,1028..1204,1365..1744)
/partial
/note="Complement factor H-related protein"
/codon_start=1
/protein_id="AAA37415.1"
/db_xref="GI:387128"
/translation="MGFCNMLLSNILLTAMLSSTAKGEKTCSPYILNGIYTPHRII
HRSDDEIRYECNIGYPVGTSTVSKCTPTGIPVPRCTLKPECFQFKYGRLYVEESL
RPFVPSIGNKYSYRCNDGFSPPSGYSDWYLRCTAQWEPVPCVRKGVFHVENGDS
AYWKIYVQGSGLKQCYNGYSLONGODIMTCTENGWSPPKCIRIKTCSADIIHIDN
GFLSESSYIYALNRTSYRCKOGYVNTGEISGTSITCLONGWSPPOPSCKSDMPVE
NSITKNTBWEKLDKLDYECLVGFENEYKHTKSGITCTIYGNWSDTPSCYDSTRTCGP
PPIDNGDITSLPEYELPSVDYQCOKYILLKGNKTTICRNGKWSPEPTCLHACVI
PENIMEAHNIIILKWHHTEKIYAHSGEDIEFECKRGYQKARGSLPFRTRCINGTNYPT
CV"
misc_feature 876..1027
/note="intron-like sequence"
exon 1028..1204
misc_feature 1205..1364
/note="intron-like sequence"
exon 1365..1744
/note="Complement factor H-related protein"
BASE COUNT 779 a 430 c 409 g 738 t
ORIGIN Chromosome 1.

Query Match 14.1%; Score 594.6; DB 10; Length 2356;
Best Local Similarity 65.2%; Pred. No. 3.4e-122;
Matches 1002; Conservative 0; Mismatches 369; Indels 166; Gaps 2;

QY 793 ctggatgaatcctcagccttctgtgaagaatgacatgtttgactccatatatccaa 852
DB 117 CTGATGCGCTTCTACTGCTAAAGGGGAGAAAGACATGCTCACCCTCTTATTTCTAA 176

QY 853 atggtatctacacctcacaggtataaacacagaaattgatgatgaatcagatatgaat 912
DB 177 ATGGTATCTACACCTCACAGGATATATACACAGAGTGATGATGAATCAGATATGAAT 236

QY 913 gtaaaaatggcttctatctcgaaccgatccctgtttcgaagtacaaattactggct 972
DB 237 GTAATATGGCTTCTATCTGTAACCTGGATGAACCTGTTTCAAGGTGTACACCCTGGCT 296

QY 973 ggatccctgctcgaagatgtagctgaacctgtgattttccacaattcacaacatggac 1032
DB 297 GATCCCTGTTCCAGAGTGTACTTGAACCTTGTGAATTTCCACAAATCAAATATGGAC 356

QY 1033 gtctgtattatgaagaagcggagaccctacttcccagtaacctataggaaggaggtaca 1092
DB 357 GTCTGTATTATGAAGAGAGCCTTGACACCACTTCCCAGTATCTATAGGAATAGTACA 416

QY 1093 gctataactgtgacaacgggtttacaacgcttcacagtcactcactcagctacettcgtt 1152
DB 417 GCTATAGGTGTGACAAACGGGTTTCCACCCTTCTGGGTATTCTCTGGGACTACCTTCGT 476

QY 1153 gcacagtaaatgggtggagcctgaagtccactgctcagcgaatgtattttccattatg 1212
DB 477 GCACAGCACAGGTTGGGAGGCTGAAGTCCCATGGCTCAGGAATGTGTTTCCATTATG 536

QY 1213 tggaaatggagaatcttcatactggaagaagatatatagaggtcagtcctgcaaaag 1272
DB 1213 tggaaatggagaatcttcatactggaagaagatatatagaggtcagtcctgcaaaag 1272

DB 537 TGGAGAAATGGAGACACTCTGCATACTGCGAAAAAATATATATGTGCGGCTCAGTCTTTAAAG 596
QY 1273 tccagtgctcagtgctctatagctctccaaatgggtcaacatcatattattgtacagaga 1332
DB 597 TCCAGTGTTCACAAATGGCTATAGTCTTCAAAATGGTCAAGACATAATGACATGTACAGAGA 656
QY 1333 atggctggctccctcctcccaaatcgctcgctatccaaagacttggttcagtatcagatatag 1392
DB 657 ATGGCTGGTCCCTCTCCCAAAATGCATCCGTATCAAGACATGTTCAGCATCAGATATAC 716
QY 1393 aaatgaaaaatgggttttttctgaatcgtgattatatacatatcgtctctaaalagaaaaaac 1452
DB 717 ACATTGACAAATGGATTCTTCTGAATCTTCTATATATATATGCTCTTAATAGAGAAACAT 776
QY 1453 ggtatagatgtaaacagggatataataacaaatcccgagaaataatcaggaaataactt 1512
DB 777 CCTATAGATGTAAAGCAGGGATATGTGACAAATACTCGAGAAATATCAGGATCAATACCT 836
QY 1513 gctctaaagatggatgggtgcacctgcacccctcatgatta----- 1551
DB 837 GCCTTCAAAATGGATGGTGCACCTCAACCTCATCGCATTTATGAAGAAGTCTTTGAAGAGCT 896
QY 1552 ----- 1551
DB 897 GCAACTCCCTGAAAGTGTGATGACCTTATAATATGTTTGACACTGTGAGTTAAGGTCC 956
QY 1552 ----- 1551
DB 957 ATGTCAGAACATCTCCCTCTTAGAAGATCTTTTAAATATTTTGATATTTGGAAGAAAGA 1016
QY 1552 -----agtctgtgatatgcctgctatttgagaattctatgactaagaataataa 1600
DB 1017 AATGCTTCCAGAGTCTTGTGATATGCTGTATTTGAGAATTTCTATTAAGAAATCTAG 1076
QY 1601 cacatggtttaaacactcaatgacaaattagactatgaatgcacattggaatgaaataa 1660
DB 1077 GACATGGTTTAAAGCTCAATGACAAATTAGACTATGATGCTCGTTGGATTGCAAAATGA 1136
QY 1661 atataacataccaaaggctctataacatgacttactatgatgagtgctctagtcacacccctc 1720
DB 1137 ATATAACATACCAAGGCTCTTAACATGTACTTATTATGATGCTGTGATACACCTCT 1196
QY 1721 ctggtatgaagaagaatgcagcattccctcgtttacaccaagacttagttgtttttccag 1780
DB 1197 ATGTTATGTTCTCCATCCCTCTACCTCCTGTGATCATTTTGTGCCCCCTTCTAAGTGG 1256
QY 1781 aagaataaaatacaaaagttggagattcgttgagtttctcttccctcagggacacagaat 1840
DB 1257 GATTGAAGCATATACACTTTTGACACTTTTGGTCTTACTTCTTTTAAAGCATCATATGGTCT 1316
QY 1841 tggagcagatttagtgcaatgctaccactttggatgggtcccttaatttcccaacggtgga 1900
DB 1317 ATGAATTGTTATGTTGGTATTTCTGAGCTTTTGGGATATCTACATATCAACTCAACAAGA 1376
QY 1901 eggccaagtataaatcatgtgaccaacctcttgaaatcccgaaatggggaaataaagggaac 1960
DB 1377 -----CATGTGGGCTCTCCACCTATTGACAAATGGAGACATCACCTCTT 1422
QY 1961 aaaaaagttgaatacacagccatggtagcgtggtggaatggaatgattgcaacctagattct 2020
DB 1423 GTCATTACGAGNATGAAACCAATTTATCATCAGTTGACTATCAGTGCCAGAAATATTATCT 1482
QY 2021 actgaagggaacccaataaaatccagttggtgacgggaagtggacaaggtttgccgatatg 2080
DB 1483 CCTTAAGGGAATAAGACAAATGAACATGTAGAATGGAAAGTGGTCTGAGCCACCAACGTG 1542
QY 2081 cgttgagtatgagagaacatgtggagaccttccctgaacttgagcagcatggtctgtcaagtt 2140
DB 1543 CTTACATGCATGTGTAATACCAAGAAACATTTATGGAAGCACACAATAATAATTTCTCAATG 1602
QY 2141 atctgtccctccctaccatcatcattgagattcagttggag 2177
DB 1603 GAGACACACTGAAAAGATTTTATGCCCCATTCAGGGGAG 1639

```

RESULT 15
LOCUS BC012610 1486 bp mRNA linear PRI 20-SEP-2001
DEFINITION Homo sapiens, clone IMAGE:3996233, mRNA.
ACCESSION BC012610
VERSION BC012610.1 GI:15706463
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1486)
Strausberg,R.
Direct Submission
Submitted (15-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Ketteman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 19 Row: a Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4504374.
FEATURES
source
1..1486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3996233"
/tissue_type="Bladder, carcinoma"
/clone_lib="NIH_MGC_53"
/lab_host="DH10B"
/notes="Vector: pDNR-LIB"

BASE COUNT 471 a 270 c 331 g 414 t
ORIGIN

Query Match 11.7%; Score 493.2; DB 9; Length 1486;
Best Local Similarity 64.1%; Pred. No. 1.4e-99;
Matches 874; Conservative 0; Mismatches 298; Indels 192; Gaps 1;

Qy 5 gtcaactgctccagatagatcccaagacatgagactgtcagcaagaattatttgcttat 64
Db 39 GTAAATGTCCTCTTAAAGATCCAAAATCAGACTTCTAGCAAGATATTTCCTTAT 98

Qy 65 attatggaactttgttagcagaagattgtcaaggctctctccagagagaattcaga 124
Db 99 GTTATGGCTATTGTGTAGCAGAAGATTGCAATCAACTTCTCTCAAGAAGATACAGA 158

Qy 125 aattctctcagggttcggtgtgaactatatcagaagggcactcagcaacctacaa 184
Db 159 AATTCTCACAGGTTCTGGTCTGACCAACATATCCAGAGAGGCCACCCAGGCTATCTATAA 218

Qy 185 atgcgcctcggataccgaacctgtgtactattgttaaaagtatgcagaagtgagaatg 244
Db 219 ATGCCGCCCTGGATATAGATCTCTGTGAAATGTAATAATGATGTCAGGAGGAGATG 278

Qy 245 ggtacctcttaaccatcaaggatattgtcggaagagccatgtgtgggcatcccgagacac 304
Db 279 GGTTCCTCTTAATCCATTAAGGAAATGTCAGAAAAGGCCCTGTGGACATCCTGGAGATAC 338

```

```

Qy 305 accotttggtcccttagctggcagtggtgactgaatttgaatttgggtgcaaaagtgt 364
Db 339 TCCTTTGGTACTTTTACCCTTACAGGAGGAAATGTGTTTGAATATGTTGTAAGAGCTGT 398

Qy 365 ttatactgtatgaagggttaccactactagggtgaataattgattaccctgtgaatgc 424
Db 399 GTATACATGTAAATGAGGGGTATCAATGCTAGGTGAGATTAATTTACCGTGAATGTGACAC 458

Qy 425 agatgggtggccaatgatattccaataatgtgaagtgtgaagtgtgtgagtgctgtccagtgacaga 484
Db 459 AGATGGATGGACCAATGATATTCCTATATGTGAA----- 492

Qy 485 actggagaatggaagaattgtgagtggtgcagccgaaccagaccaggaatatattatttgg 544
Db 493 ----- 492

Qy 545 acaggtgtacgtttgaatgcaactccggttccaagattgaaggacagagaagaatgca 604
Db 493 ----- 492

Qy 605 ctgctcataaaatggcctctggagcaatgaaaagccacagtggtggaaatttcttgcct 664
Db 493 -----GAATTTTCATGCCAA 506

Qy 665 gccaccacgagttgaaaatggagatggtatatctgtgaaccagtttacaaggagaatga 724
Db 507 ATCCCCAGATGTTATAATGATGATCTCCTATATCTCAGAAGATTTATTTAAGGAGAGATGA 566

Qy 725 aagattccaataataatgaagcaggtttgtgtacaaaagaagagggtgtgtgtctg 784
Db 567 ACGATTTCAATATAATGTAACATGGGTATGAATACAGTGAAGAGGAGATGCTGTATG 626

Qy 785 cacgggttctggatggaatccctcagccttctgtgaaagaaatgacatgtttgactcccata 844
Db 627 CACTGAATCTGGATGGCGTCCGTTGCTTCATGTGAAGAAAAATCATGTGATAATCCTTA 686

Qy 845 tatccaaaatggttatctcacacacctcacaggattataaacacagaattgtagatgaatcag 904
Db 687 TATTCCAAATGGTGACTACTCACCTTTAAGGATTTAAACACAGAACTGGAGATGAATCAC 746

Qy 905 atatgaatgtaaaaatggctctctatccctgcaccccgatcacctgtttcaaaagtacaa 964
Db 747 GTACCAGGTAGAAATGGTTTTATCTCTGCAACCCGGGGAAATACAGCCAAATGCACAAG 806

Qy 965 tactggctggaatccctgtcccaagatgtagcttgaacacctgtgtatttccacaattcaa 1024
Db 807 TACTGGCTGGATACCTGCTCGAGATGTACCTTGAACCTTGTGATTATCCAGACATTAA 866

Qy 1025 acatggacgtctgtattatgaagaaagccggagagaccctacttcccagtaacctataggaaa 1084
Db 867 ACATGGGAGGTCTATATCATGAGAAATATGCGTAGACCATACTTTCCAGTAGCTGTAGGAAA 926

Qy 1085 ggagtacagctataactgtgacaacgggtttcaacgccttcacagtcatactgggacta 1144
Db 927 ATATTACTCTTATTACTGTGTATGAACATTTTGAGACTCCGTCAGGAAGTTACTGGGATCA 986

Qy 1145 ccttcgttgacagataaatgggtgggagcctgaagtccatgcctcagggcaattatttt 1204
Db 987 CATTCATTGTCACACAAGATGGATGGTGCAGCAGCATACCATGCTCTCAGAAAAATGTTATT 1046

Qy 1205 ccattatgtggaatatgagagaattctcatactggcgaagaagatatatagaggggtcagtc 1264
Db 1047 TCCTTATTTGGAAAAATGATATAATAAATTTATGGAAGAAAGTTTGTACAGGGTAAATC 1106

Qy 1265 tgcaaaagtcagtgctcacagtggtcattagcttccaaaatggttcaagatacatattattg 1324
Db 1107 TATAGAGTTTGGCTGCCATCCTGTGCTAGCTCTTCCAAAAGCCGACAGACAGTACATG 1166

Qy 1325 tacagagaatggctgtgtccctccctcccaaatgcgtccglatca 1368
Db 1167 TATGGAGAATGGCTGGTCTCCTACTCCAGATGATCCGTTCA 1210

```

Search completed: August 30, 2002, 19:06:49
Job time: 11894 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 14:57:15 ; Search time 3911.83 Seconds
(without alignments)
14591.284 Million cell updates/sec

Title: US-09-316-163-1
Perfect score: 4229
Sequence: 1 tcagatcaactgctcccgag.....aaaaaaaaaaaaaaaaaaaaa 4229

Scoring table: IDENTITY_NUC
Gapop 10.0 , Capext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estto: *
8: em_htc: *
9: gb_esti: *
10: gb_est2: *
11: gb_htc: *
12: gb_gss: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vrt: *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	671.8	15.9	1113	10	BM461070
2	583.4	13.8	881	9	AA833130
3	558.8	13.2	903	9	AL531827
4	558.2	13.2	818	9	AI006101
5	554.6	13.1	906	9	AL540473
6	544.6	12.9	631	9	AI170314
7	538	12.7	626	9	AI009773
8	534	12.6	726	10	BI327605
9	520.6	12.3	548	10	BM383739
10	517	12.2	665	9	BB617140
11	514.4	12.2	661	9	AW701301
12	512.2	12.1	652	10	BF168914
13	510.6	12.1	842	9	AI122695
14	505.2	11.9	759	9	AA882470
15	496.6	11.7	652	9	BB609000
16	494.4	11.7	891	10	BF237071
17	494	11.7	595	10	BF395144

ALIGNMENTS

RESULT 1

BM461070 1113 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6419764 NCI_CGAP_OV44 Mus musculus cDNA clone
DEFINITION IMAGE:5503808 5', mRNA sequence.
ACCESSION BM461070
VERSION BM461070.1 GI:18510110
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1113)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Aaron Hsueh
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM2144 row: c column: 09
High quality sequence stop: 662.
Location/Qualifiers
1. .1113
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5503808"
/clone_lib="NCI_CGAP_OV44"
/lab_host="DHI08 (T1 phage-resistant)"
/note="Organ: ovary, PMSG-treated; Vector: pCMV-SPORT6.cdb; Site: 1: EcoRV; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2-2 kb. Library constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 353 a 224 c 235 g 300 t 1 others

Db	976	GGCCTGAATCAACTGCGTTTAAACCCATTAAACCCCAATTTACTAGATTTTACCCCTAAC	1035
QY	2352	tttagtgtagttacagatagacaaaagcagagtagatgaacatccaatcgtcatcaat	2411
Db	1036	TCCCCCTGGATTACCAAGTGT--GAACAGCGGGGATACCAACGCTCA--CCTGGATCAG	1091
QY	2412	ggaagatggggtcctgaaacaa	2433
Db	1092	GGCAATGGGATCCTGAACCAA	1113
RESULT	2		
AA833130			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
Source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	1351	ccaaatgctccgtatcaagaactgttcagtatcagatatagaaattgaaatgggtttt	1410
Db	17	CCCAATGCATCCGTATCAACAGATGTTTCAGCATGATACACATACACATGATTC	76
QY	1411	tttctgaatcgtattacatcatgctctctaaatagaaaaaacacggtatagatgtaaacag	1470
Db	77	TTTCTGAATCTCTCTATATATAGTCTAAATAGAGAAACATCCTATAGATGTAGACAG	136
QY	1471	gatattgaaacaataacc-ggagaaatatcaggaataattacttcttccaagatggatgg	1529
Db	137	GATATGTGACAAATACNTGAGAAATATCAGGATCAATAAAGTTCCTTCAAAATGGATGG	196
QY	1530	tcacctgagacctcatgacttaagttctggtatgacctgtattgttgagaattctatgact	1589
Db	197	TCACCTCAACCCCTCATGCAATTAAGTCTTGTGATATGCCGTATTTGAGAATTTCTATAACT	256
QY	1590	aagaataataacacatggtttaaacctcaatgacaattagacatgaaatgacatggaa	1649
Db	257	ARGAATACTAGGACATGGTTTAAGCTCAATGACAAATTAGACTATGATGCTCGTTGGA	316
QY	1650	tatgaaatgaatatataaacatccaaaggctctataacatgtaacttatggtggatgct	1709
Db	317	TTTGAATAATGAATATAAACATACCAAAAGCTCTATAACATGTACTTATTATGGATGGTCT	376
QY	1710	agtaacctctgttatgagaagaatgcagacatccctctgtttacacaaagactlaagt	1769
Db	377	GATACACCCCTCATGTTTATGAAGAGAAATGCAGTGTGCCACTCTAGACCGAAACATGCTC	436
QY	1770	gttttccagagaagttaaatacaaaagttggagatcgttgagttctcttgcctgtca	1829
Db	437	GTTTCCCCCAGAAAAGAAATACAGAGTTGGAGATTTGTTGGAATTTCTCTCGCATTTCA	496
QY	1830	ggacacagagttggagcagatttagtgcgaatgctaccactttggatggtccctcaatttc	1889
Db	497	GGACACAGAGTTGGGCGCAGATTTCAGTGCATGCTACCCTTTGGATGCTCCTCGGTTTC	556
QY	1890	cgaactgtg---gaagggcaagtaaaatcatatgaccacacctcttgaatacccaaatggg	1946
Db	557	CTACATATGTAACACAGGCTCAAGTAGCATCATGTGCACCCACTTTGAAATTTCTTAATGGG	616
QY	1947	gaataaagggaacaaaagattgaatacacaccatgtagcgtgtagaataatgattgc	2006
Db	617	GAAATTAATGAGCAAAAAGTTGAATACACCCATGGTGAAGTGGTGAATATGATTTC	676
QY	2007	aaacctagatttctactgaagggaccacataaaatccagtggttgacgggaagtggaca	2066
Db	677	AAACCTAGATTCTACTGAAGGAGGCCCAATAAAATCCAGTGTGTGATGGGAATTTGGACA	736
QY	2067	agtttccagatatgcgttagtatgagaacacatgtagaaccttccctgaacttgagcat	2126
Db	737	ACCTTGCCCTGTATGTATTGAGGAGGAGGAAC-TGTGGAGACATTCCTTGAACCTTGAACAT	795
QY	2127	ggctcgtcaagt-tatctgctccctccctaccatcatgagatcagtgaggttcaacttg	2185
Db	796	GGCTCTGCCAAGGTTCTCTGTTCCCTTACCACCATGGAGATTCAGTGGGAAGTCATTTC	855
QY	2186	tacagaaaccttcacaaatgattggacat-gcagtagttttctgcaattagtggaaggtgga	2244
Db	856	TGAAGAAACTTCACATGAATTTGGACATGGGTCCAGTTTCTTGCAATAGTCGAAATTTGGA	915
QY	2245	ccagatttctcaatggtgtgca-----acagatcaactggagaaagtgt-aaa	2291
Db	916	CCCAGTTTCTTAAAGGGGTTGATTTTCTTATGACACAGACACCCTTGGAAGTGTAAAA	975
QY	2292	gccccgaatcaactggaatagatgaacttcaataagaatgaatttaatacataaac	2351

AUTHORS Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1..906
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DE001YB21"
 /clone_lib="LRI FL002_PL1"
 /lab_host="DH10B"
 /note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 299 a 138 c 214 g 255 t
 ORIGIN

Query Match 13.1%; Score 554.6; DB 9; Length 906;
 Best Local Similarity 77.8%; Pred. No. 6.e-88;
 Matches 682; Conservative 0; Mismatches 194; Indels 1; Gaps 1;

Qy 5 gTcaactgtccagatagatccagacatgagactgtcagcaagaattattgtcttat 64
 Db 30 GTAATGTCCTCTTAAAGATCAAAAATGAGACTTCTAGCAAGGATATTGTGCTTAT 89
 Qy 65 attatggactgttgttagcagaagattgtaaaggtctctcccaagagaattcaga 124
 Db 90 GTATGGGCTATTGTGAGCAGAGATTGCAATGAACCTTCTCCAGAGAAATACAGA 149
 Qy 125 aattctctcagttcgtgtgtgtgaacaactattcagaaggcactcaggcaacctcaaa 184
 Db 150 AATTCTGACAGGTTCTGTGCTGTGACCAACATATCCAGAAAGCCACCCAGGCTATCTATAA 209
 Qy 185 atgcgcctgtgatacccaaacacttggtactatttataaagtatcagaagaatggagaatg 244
 Db 210 ATGCCGCCCTGGATATAGATCTCTTGGAAATGTAATAATGGTATGATGACGAGGACAATG 269
 Qy 245 ggtactcttaaccatcaaggatattgtcggaaaaggccatgtggcattcccgagagacac 304
 Db 270 GGTGCTCTTATCCATTAAAGGAATGTGAGAAAGGCCCTGTGCACATCTCGGAGATAC 329
 Qy 305 acccttgggtcctttagctggcagtggtgatctgaaatttggaatttggtgcaaaagtgtt 364
 Db 330 TCCTTTTGTACTTTTACCCTTACAGGAGGAATCTGTTTGAATATGTTGTTAAAGCTGT 389
 Qy 365 ttatacatgtgatgaagggtaccactattagtgaaattgattccctggaatgtgagtc 424
 Db 390 GTATACATGTAATAGGGGTATCAATTCTAGGTGAGATTAATTACCGTGAATGTGACAC 449
 Qy 425 agatgggtggacccatgatattcccaatattgaaattgtgaagtcttggaagtcttgccagtacaga 484
 Db 450 AGATGGATGGACCAATGATATTCCTATATGTGAAAGTTGTGAAGTTTACCAAGTGACAGC 509
 Qy 485 actggagaatggaagaattgtgagtggtgagccgcaaacacagaccaggaataattatttgg 544
 Db 510 ACCAGAGAATGAAAATTTGTGAGTGTGCAATGCAATGCAACAGATCGGGAATACCATTTTGG 569
 Qy 545 acaggtgtacgcttgaatccaactccggttcaagattgaaggacagaagaataatgca 604
 Db 570 ACAAGCAGTACGGTTGTATGTAATTCAGGCTACAGGATGAAGGAGATGAAGAAATGCA 629
 Qy 605 ctgctcaaaaatggccctctggagcaatgaaaagccacagtggtgtggaaattcttgcct 664

Db 630 TTGTTTCAGACGATGGTTTTTGGAGTAAAGAGAAACCAAAAGTGTGGAATTTTCATGCCAA 689
 Qy 665 gcccaccagagttgaaatggagatggtatatctgataaccagttttacaagagaatga 724
 Db 690 ATCCCCAGATGTTATAAATGGATCTCTATATCTCAGAAGATTATTATTATGAGGAGAAATGA 749
 Qy 725 aagattcccaataataatgtaagaaggtttttgtgtacaaagaaggaggtgctgtctg 784
 Db 750 AGCATTTCAATATAAATGTAACATGGTTTATGAATACAGTGAAGAGGAGAGATCTCTATG 809
 Qy 785 caagggttctggatgaaatccctcagccttccctgtgagaagaatgacatgtttg-actccat 843
 Db 810 CACTGAATCTGGATGGCGTTCCTTCATGTTGAAGAAAATCATCTGTCAGTAAATCCCTT 869
 Qy 844 atattcccaatggtattctacacacccacagagattaa 880
 Db 870 ATATTCCAATGGTGACTACTCACCTTTAAGGATTAA 906

RESULT 6
 AI170314/c
 LOCUS AI170314 631 bp mRNA linear EST 20-JAN-1999
 DEFINITION EST216240 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
 RLU0G82 3' end, mRNA sequence.
 ACCESSION AI170314
 VERSION AI170314.1 GI:3710354
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 631)
 AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 TITLE Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
 Gene Index
 JOURNAL Unpublished (1998)
 COMMENT Other_ESTs: TC52298
 Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.

FEATURES
 source
 1..631
 /organism="Rattus sp."
 /db_xref="ATCC (inhost):2027791"
 /db_xref="taxon:10118"
 /clone="RLU0G82"
 /clone_lib="Normalized rat lung, Bento Soares"
 /note="Organ: lung; Vector: pT73Pac; Site_1: EcoRI;
 Site_2: NotI"

BASE COUNT 212 a 97 c 104 g 218 t
 ORIGIN

Query Match 12.9%; Score 544.6; DB 9; Length 631;
 Best Local Similarity 96.0%; Pred. No. 4e-86;
 Matches 606; Conservative 0; Mismatches 9; Indels 16; Gaps 4;

Qy 3588 ctccagatgaggaataatgcaagaatttattcccaatcaggagagaattgaattcatg 3647
 Db 631 CTCAGATGAGGGAATAATGCAAGATTTATCCCAATCAGGGGAGAAATTTGAATTCATG 572
 Qy 3648 tgaacacctggatagaaaaattcagagagatcacctccgttttcgtaaacagtgcatgag 3707
 Db 571 TGTAAACCTGGATATAGAAAATTCAGAGGATCACCTCCGTTTCGTACAAAGTGCATTGAG 512
 Qy 3708 ggtccatcaaatcccccacttggtataaaaatcgatatcaaatattagtaaaccttat 3767

```
|||||
Db 511 GGTACATCAATTATCCACTGTGTATAAAATCGCTATACAATATTAGTAAACCTTAT 452
Qy 3768 ggaag------aqaatgcacatgtatattactaataacacagatttgaatttaccatt 3816
Db 451 GGTAAACCTTTGTTAGAAATGACATGTATATTACTTAATACAGTTTGAATTTTACATTT 392
Qy 3817 -aaatatgttttagctcttctcttaataagatatataaaactttttttatatlggtgglta 3875
Db 391 GAAATATTGTTTAGCTCATTTCTCTAATAAGTATATAAATTTTATATATGTTGGTTA 332
Qy 3876 atcagtaactttacagactgttgcacaaagcaagaacattacattcacaactcctaatc 3935
Db 331 ATCAGTAACTTTACAGACTGTGTCCACAAAGCAGAACATTTGCATTTCAAAACTCCTAATC 272
Qy 3936 c-aaatatgatgtccaaagacaaactatgtctaaagcaagaaataaataatttaattt 3994
Db 271 CAAATATGATATGTTCCAAAGCAACAATATGTCTAAGCAAGAAATAAATGTTAGTCTT 212
Qy 3995 caatgtctgtttttatttcaggacacctttcagattttcttggataacct---ttgttaggtt 4051
Db 211 CAATGCTGTTTTTATTTCAGGACTTTTTCAGATTTTCTTGGATACCTTTTGTGTTAGGTT 152
Qy 4052 ctgattcacagtgaatgaagacacactgaactgaacttcaaataggattacttgcgaat 4111
Db 151 CGATTTCAGTGTGAGGAGACACACTGACTCTGACTTCAAAATGATTACTTTGCCAA 92
Qy 4112 acattaaacaaactatcataatcacaaatgtatcacagctaatctactgtgtcctac 4171
Db 91 TACATAACAACAACTATCATATATATACAAATGTATACAGCTAATTTACTGTGCTCTAC 32
Qy 4172 cttgtatcaataaagaatactagaagaagtt 4202
Db 31 CTTTGTATCAATAAGAAATCTAAGAAAGTT 1

RESULT 7
LOCUS AI009773/c 626 bp mRNA linear EST 25-JAN-1999
DEFINITION EST204224 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
ACCESSION AI009773
VERSION AI009773.1 GI:3223605
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..626
/organism="Rattus sp."
/db_xref="ATCC (inhost):2020517"
/db_xref="taxon:10118"
/clone="RLJB029"
/clone.lib="Normalized rat lung, Bento Soares"
/notes="Organ: lung; Vector: pT73pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 213 a 95 c 103 g 215 t
ORIGIN
```

Query Match 12.7%; Score 538; DB 9; Length 626;

Best Local Similarity 95.8%; Pred.No.5.9e-85;

Matches 600; Conservative 0; Mismatches 10; Indels 16; Gaps 4;

Qy 3571 aaaaacataataagttctcagatggagggaataatgcaagaatttattcccaatcagg99 3630

Db 626 AAAAAACATAATAGTCTCAGATGGAGGGAATAATGCAAGATTATTCCCAATCAGGGG 567

Qy 3631 agaattatgaattcattgtgttaaaccttgatataagaaaattcagagatcacctccglttc 3690

Db 566 AGAATATTGAATTCATGTGTAAACCTGGATATAGAAATTCAGAGGATCACCTCCGTTTC 507

Qy 3691 gtacaaagtgaatgagggtcacatcaattatcccaacttgttataaaaactcactacaa 3750

Db 506 GTCAAAAGTGCAATGAGGGTCACATCAATTTATCCCACTGTGTATAAAATTCGCTATACAA 447

Qy 3751 ttattagtaaaccttatggatg-----agaaatgcacatgtatatlactaataac 3799

Db 446 TTATTAGTAAACCTTATGGATGAACCTTTGTTTAGAAATGCACATGTATATTACTTAATAC 387

Qy 3800 agtttgaatttaccattt-aaatatgttttagctcttctctctaataagtatataaactt 3858

Db 386 AGTTTGAATTTACATTTGAAATATTGTTTAGCTCATTTCTCTAATAAGTATATAAATTT 327

Qy 3859 tttttatatgtgtggttaataatcagtaactttacagactgttgcacaaagcaagaacattac 3918

Db 326 TTTTATATGTTGGTTTATCAGTAACCTTTACAGACTGTGGCCACAAACCAAGCAATGCC 267

Qy 3919 attcaaaactcctaatcc-aaatatgatattgtccaaagacaaactatgtctaaagcaaga 3977

Db 266 ATTCAAAACTCCTAATCCAAATATGATATGTCCAAGGCAAACTATGTCTAAGCAAGAA 207

Qy 3978 aataaatttagttcttccaatgtctgtttttatttcaggacaccttcagattttcttggata 4037

Db 206 AATAAATGTTAGTCTCTCAATGTCGTTTTATTTCAGGACTTTTCAGATTTTCTTGGATA 147

Qy 4038 cctt---ttgttagttcttgcacagtgagtggaagacacactgactctgaacttcaaa 4094

Db 146 CCTTTGTTGTTAGGTTTGTGATTTCACAGTGAGTGGGAAGACACACTGCTGACTTCAAA 87

Qy 4095 ttagtattacttgcaatatacatataacacaaactatcatataatcacaaatgtatcacgc 4154

Db 86 TTAGTATTACTTGCCAATACATACAAACCAAACTATCATATATATCAAAATGTATACAGC 27

Qy 4155 taattactgtctcctaccttggatc 4180

Db 26 TAATTACTGTGCTCTACCTTTTGTATC 1

RESULT 8

LOCUS BI327605

DEFINITION BI327605

ACCESSION BI327605

VERSION BI327605.1

KEYWORDS GI:15012262

SOURCE EST.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NIH-MGC http://mgc.nci.nih.gov/.

1 (bases 1 to 726)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.

LOCATION/QUALITY
1. .726

BASE COUNT
ORIGIN

[illegible]

QY	1899	---	gaagggccagtaataatcatgtgacacacactcttgaatcccgaaatcggggaataaag	1955
Db	242	AAAGCAGGTCAAGTAGCATCATGTGCACAACTCTGAATTTCTTAATGGGGAATTAAT	301	
QY	1956	ggaacacaaaaagttagaacacacacatgtagctggtggaatgatgcacaaacctaga	2015	
Db	302	GGAGCAAAAAGAGTGTGAATACAGCCCATGUGTAGGTGGGTATGATGTCGAACCTAGA	361	

	2075	t t t c t a c t g a a g g g a c c c a a t a a a t c c a g t g t t g a c g g a g t g a c a a g g t t a c c y
Qy	2016	t t t c t a c t g a a g g g a c c c a a t a a a t c c a g t g t t g a c g g a g t g a c a a g g t t a c c y
Db	362	T T C C T A C T G A A G G G A C C C A A T A A A A T C C A G T G T G T T G A T G G A T G T G G A C A A C C T T G S C C T
	421	a t a t c g t t g a g t a t g a g a a c a t g t g a g a c c t c t c t a a c t t g a g a c t g a c t c t g t c
Qy	2076	a t a t c g t t g a g t a t g a g a a c a t g t g a g a c c t c t c t a a c t t g a g a c t g a c t c t g t c
Db	422	G T A T G T G T T G A G A G G A G A A C A T G T G G A C A C A T T C C T G A A C T T G A A C A T G C C T C T G T C
	481	

	Oy	2196	ttcacaatgattggacatgcagtagtctttctcgcattagtggaaaggtagcccgagcttccct	2255
	Db	542	ttcacaaatgattggactttgggtcagtttcttgcccttaagtggaaagtgagccccagctttccct	601
	Oy	2197	gaatttatctgtgccctccctaccatacatgatgattcagtgaggttcacctgtcagaataacc	2199
	Db	482	aagttttgtctcccctccctaccacatggagatttcagtgaggtttcacttctgcagaataacc	541

602 AATGTTGTCACAGACCAACTGGAGAAGTGTAGTGTGAAGTCAACTGCATAGAA 661
 Db
 2316 gcaattccataaaga--tgaaattaatcataactttagtgtgagttcacagatga 2373
 QY
 662 GCAATTAAACCAAAAACAACTGAATTTTCAGCATAACTCCAGCATGTATTACAAATGTC 721
 Db

RESULT 9
BM383739/C

REFERENCE AUTHORS TITLE	JOURNAL MEDLINE COMMENT
1. KIMURA T, KAWAKATSU M, YAMAGUCHI S, et al. The effect of the duration of treatment with cyclosporin A on the expression of HLA antigens in renal allograft recipients.	
2. KIMURA T, KAWAKATSU M, YAMAGUCHI S, et al. Effect of cyclosporin A on the expression of HLA antigens in renal allograft recipients.	
3. KIMURA T, KAWAKATSU M, YAMAGUCHI S, et al. Effect of cyclosporin A on the expression of HLA antigens in renal allograft recipients.	
4. KIMURA T, KAWAKATSU M, YAMAGUCHI S, et al. Effect of cyclosporin A on the expression of HLA antigens in renal allograft recipients.	
5. KIMURA T, KAWAKATSU M, YAMAGUCHI S, et al. Effect of cyclosporin A on the expression of HLA antigens in renal allograft recipients.	
6. KIMURA T, KAWAKATSU M, YAMAGUCHI S, et al. Effect of cyclosporin A on the expression of HLA antigens in renal allograft recipients.	
7. KIMURA T, KAWAKATSU M, YAMAGUCHI S, et al. Effect of cyclosporin A on the expression of HLA antigens in renal allograft recipients.	
8. KIMURA T, KAWAKATSU M, YAMAGUCHI S, et al. Effect of cyclosporin A on the expression of HLA antigens in renal allograft recipients.	
9. KIMURA T, KAWAKATSU M, YAMAGUCHI S, et al. Effect of cyclosporin A on the expression of HLA antigens in renal allograft recipients.	
10. KIMURA T, KAWAKATSU M, YAMAGUCHI S, et al. Effect of cyclosporin A on the expression of HLA antigens in renal allograft recipients.	

[illegible]

BASE COUNT	141 a	123 c	106 g	178 t
ORIGIN				

Query Match 12.3%; Score 520.6; DB 10; Length 548;
Best Local Similarity 97.8%; Pred. No. 7e-82;
Matches 539; Conservative 0; Mismatches 9; Indels 3.

Qy 2515 atggagaaaagtatctgttcttgcgaagatggttacctaaactcaaggqccccagaaqaaa 2574


```

|||||
Db 242 CAATGAGACATCCCTCTCTCATTACCATGATATGACCATATATCATCATGTCGAATA 301
QY 3455 tcaatgccagaactattatctacttaagggaaataagatagtaacatgtagaataaggaaa 3514
Db 302 TCAATGCCAGAAGTATTATCTCTTAAAGGGAAGAAGACAAATACATGTAGAAATGGAAA 361
QY 3515 gtggctcagccaccacactgtctacatgcacatgcatgtgtgataccagagaagatatatggaaaa 3574
Db 362 GTGGCTGAGCCACCACCAATGCTTACATGCATGTGTAAATACCCAGAAACATATATGGAATC 421
QY 3575 acataatagttctcagatgaggggaaataacaaagattttatcccaatcaggggagaa 3634
Db 422 ACACAATATAATTCCTCAATGAGACACACTGAAAGAGATTATTCCTCCATTCAGGGAGGA 481
QY 3635 tattgaattcattgtataacacctggatatagaaaaattcagagagatcacacctcccttctgtac 3694
Db 482 TATTCAATTTGGATGTAATATGATATGATATATAAGCAAGAGAGATTCACCCGCCATTTGCTAC 541
QY 3695 aaagtgcattgaggtcacatcaattatcccaactgtgtataaaatcgctatacaattat 3754
Db 542 AAAGTGCATTATGCGACCATCAATATATCCCACTTGTGTATAAAATPCATATACATTTAT 601
QY 3755 tagtaacctattgtagagaaatgcacatgtattactataacacagtttggatttacct 3814
Db 602 TAGTTGATTNTATGCTTAGAAAGCACACATGATGCTGACTATATACTNTCAATTTCAT 661
QY 3815 tta 3817
Db 662 TGA 664

```

```

RESULT 11
AW701301/c
LOCUS
DEFINITION
IMAGE:2537283 3' similar to gb:M12660 Mouse CFH locus, complement
protein H gene, complete cds, (MOUSE);, mRNA sequence.
ACCESSION
AW701301
VERSION
AW701301.1 GI:7585430
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 661)
AUTHORS
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Willson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
CONTACT: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:019879
Seq primer: custom primer used
High quality sequence stop: 501.
Location/Qualifiers
1..661
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2537283"
/clone_lib="Sugano mouse kidney mkia"
/sex="female"
/dev_stage="adult"

```

FEATURES

```

source
1..661
/lab_host="DH108"
/Note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
(CACTGTGG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCTTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCTCTCTTAAAGCTGGC and 3' end
primer CGACTCGACCTCGACACA."
BASE COUNT 191 a 136 c 127 g 207 t
ORIGIN
Query Match 12.2%; Score 514.4; DB 9; Length 661;
Best Local Similarity 86.2%; Pred. No. 8.3e-81;
Matches 569; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 1099 actgtgacaacgggtttacaacgcttcacagtcacactgagggactacacctcgttgcacag 1158
Db 661 AGTGTGACAAACGGGTTTTCACCACATTTCTGGGTATTTTCATGAGACTACCTTCGTGTGCACAG 602
QY 1159 taaatgggtggagcctgaagttccatgcctcagggcaatgtatttccattatgtggaat 1218
Db 601 CACAAGGGTGGGAGCCTGAAAGTCCCATGCTCAGGAAATGTGTTTCCATTATGTGGAGA 542
QY 1219 atggagaatcttcatactactggcgaagaagatatataagagggtcagctgcaaaagtcagct 1278
Db 541 ATGGAGACTTCGCATCTACTCTGGAAGATGCATATGTGCGAGGTTCAGTCTTTAAAGATCCAGT 482
QY 1279 gtcaacagtggtctatagctctccaaatggccaagatacatattatttgcagagaatggct 1338
Db 481 GTTACAATGGCTGTAGTCTTCAAAATGGTCAAGACACAATGACATGTACAGAGAAATGGCT 422
QY 1339 ggtccctcctcccaaatgcgtccgtatcaagactgttccagtatcagatatagaaattg 1398
Db 421 GGTCCCTCTCTCCCAATGCATCCGTATCAAGACATGTTTCAGCATCAGATATACACATTG 362
QY 1399 aaaaatgggtttttctgaatcgtgattacatatatgctctaaataaagaaacacggtata 1458
Db 361 ACAATGGATTTCTTCTGAATCTTCTTATATATGCTCTAAATAGAGAAACATCTCTATA 302
QY 1459 gatgtaaacagggatattgtacaataacccggagaaatatcaggaaataattactgtcttc 1518
Db 301 GATGTAAAGCAGGGATATGTGACAAATACTGGAGAAATATCAGGATCAATAACTTGCCTTC 242
QY 1519 aagatggatggttcacctgcacctcctcattgaagctctgtgatatgcctgtatttgaga 1578
Db 241 AAAATGGATGGTTCAGCTCAACCTCTCATGATTTAAGTCTTGTGATATGCTGTATTTGAGA 182
QY 1579 attctactaactaagaataaacacatggttttaactcaatcaataactagactatgaat 1638
Db 181 ATCTATTAATTAAGAATACTAGGACATGGTTTAAAGTCAATGACAAATAGACTATGAAT 122
QY 1639 gtcaattggatgataaataatgaataaataacacaaaggtctctaaactgactatg 1698
Db 121 GTCTCGTGGATTTGAAATGAATATAAACATACCAGAGCTCTATAACATGCTACTATT 62
QY 1699 atggatggtctagtcacacctcctctgtatgaagaagaatgcagcattccctgttacacc 1758
Db 61 ATGGATGTTTGTATACACCCCTCATGTTATGAAAGAGAAATGCAGTGTGCCACTCTAGCCC 2

```

RESULT 12

BF168914

LOCUS

DEFINITION

ACCESSION

```

601775377F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:4017115 5',
mRNA sequence.
BF168914
652 bp mRNA linear EST 30-OCT-2000

```

VERSION	BF168914.1	GI:11049266
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 652)	
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Inyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM9265 row: 1 column: 20 High quality sequence stop: 648.	
FEATURES	Location/Qualifiers	
source	1..652	
	/organism="Mus musculus"	
	/strain="C57BL/6J II (feral)"	
	/db_xref="taxon:10090"	
	/clone="IMAGE:4017115"	
	/clone_lib="NCI CGAP Lu29"	
	/tissue_type="spontaneous tumor, metastatic to mammary."	
	Stem cell origin"	
	/db_host="DB10B"	
	/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"	
BASE COUNT	216 a 139 c 134 g 161 t	2 others
ORIGIN		
Query Match	12.1%	Score 512.2; DB 10; Length 652;
Best Local Similarity	87.6%	Pred. No. 2e-80;
Matches 570; Conservative	0; Mismatches 80; Indels 1; Gaps 1;	
Qy 2414	aaatggggtcctgaaccacaaactgtacaagcaaaagattctgcctctctcccccgcagat 2473	
Db		
1	AAATGGGATCTCTGAACCAAACTGTACAAGCAAAACATCTCGCCTCTCTCCACGCAGAT 60	
Qy 2474	tccaaatgccaaagtattgaaccacccgtgaaataacttgatggagagaaaaagtatcgt 2533	
Db		
61	TCCAATACCCAACTGATTTAAACCCACCCTGAATACTTGGATCGAGAAAAATATCTGT 120	
Qy 2534	tccttgccaaagtatggttaactaactcagggccagaagaagaatggtgtgtaaacatggaag 2593	
Db		
121	TCCTTGCCAAAGACAATTTACCTTAACTCAGAGACTTCAGAGAAATGGTGTCAAAAGATGCAAG 180	
Qy 2594	gtggcagtcgttacacgcctgcacggagaaaaattccatgtttccagcccccataaattga 2653	
Db		
181	GTGGCAGTCATTACCTCGCTGGATTTGAAAAAATTCATGTGTTCCAGCCCCCTACAAATAGA 240	
Qy 2654	acatgatctattaaatgcgccaggtctctcagaagaggaggagatttaattgagtcacg 2713	
Db		
241	ACATGGATCTATTAATTTACCCAGACTCTTCAGAGAAAAAGAGATCCATTGAGTCCAG 300	
Qy 2714	cagttatgaacacggaactacattcagctattgtgttagagatgagattcagaatatatga 2773	
Db		
301	CAGTCATGAACATGGAACATACATTCAGCTATGCTCTGTGATGATGGTTTCAGGATACCTGA 360	
Qy 2774	agaaatagggttaacctgcacaatgggaaaaattggagctctctgcctcgttgtgttggaat 2833	
Db		
361	AGAAAAATAGGAATACCTGCTACATGGGAAATGGAGCACATCCACCTCGCTGTGTGGACT 420	
Qy 2834	acccttggagcccccaacttcaattccctcttggttatgtttctcatgaaatgaaagtta 2893	
Db		

```

|||||
226 ATGCCGCCCTGGATATAGATCTCTTGGAAATATAATAATGTTATCGAGGAGGAGATG 285
QY 245 ggtaccttctaaaccatcaaggatgtcgaaaagccatgtggccatcccgagagacac 304
Db 286 GGTTCCTCTTAATCCATTAAGGAAATGTCAGAAAAGCCCTTGTGGACATCCTGGAGATAC 345
QY 305 acccttgggtcctcttaaggctggcagttggatctgaaattgaaatttggtgcaagggtgt 364
Db 346 TCCTTTTGGTACCTTTTACCCCTTACAGGAGGAAATGTGTTTGAATATGTTGTAAGAGCTGT 405
QY 365 ttatacatgtgatgaaggggtaccacactattagggtgaaattgattaccgtgaatgtgatgc 424
Db 406 GTATACATGTAAATGAGGGGTATCAATTTGCTAGGTGAGATTAATACCGTGAATGTGNACAC 465
QY 425 agatgggtggaaccaatgatattccaatatgtgaagtgtgaaagtgttgcctgacagaga 484
Db 466 AGATGGATGGACCAATCATATTCTTATGTGAAGTGTGAAGTGTGAAGTGTGTACCAAGTGCAGC 525
QY 485 actggagaatggagaatgtgagtggtgcagccgaaccagaccagaaatattattttgg 544
Db 526 ACCAGAGAATGGAAAATGTCAGTAGTGCATGGAACCAACCAAGTGTGTGCAAAATTTTCATGCAA 585
QY 545 acaggtgtgtacgcttgaatgcaactccggcttcaagattgaaggacagaaaatgtca 604
Db 586 ACAAGCAGTACGGTCTTCTATGTAACTCAGGCTACAGATTTGAAGGAGATGAAGAAATGCA 645
QY 605 ctgctcataaaatggcctctggagcaatgaaagccacagctgtgtgaaattcttgcct 664
Db 646 TTGTTTCAGACGATGGTTTTTGGAGTAAAGAAACCAACCAAGTGTGTGCAAAATTTTCATGCAA 705
QY 665 gccaccagagttgaaatggagatggtatatatctgaaaccagtttcaagagagaatga 724
Db 706 ATCCCGAGATGTTAATAGGATCTCTATATCTCAGAGATTTATTAAGGAGAATGA 765
QY 725 aegattccaatataatgaagcaaggttttgtgtacaaagaaggagggtgctgtctgt 784
Db 766 ACGATTTCAATATAAATGTAACATGGGTATGAATACNGTGAAGAGAGANATCCTGNATG 825
QY 785 caccggttctgg 796
Db 826 CCTGGATCTGG 837

RESULT 14
LOCUS AA882470
DEFINITION vx45b10.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone
IMAGE:1278139 5' similar to gb:Y00716 COMPLEMENT FACTOR H PRECURSOR
(HUMAN); gb:M12660 Mouse CFH locus, complement protein H gene,
complete cds, (MOUSE);, mRNA sequence.
ACCESSION AA882470
VERSION AA882470.1 GI:2991581
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 759)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

```

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:669939
 Seq primer: -28ml3 rev1 ET from Amersham
 High quality sequence stop: 439.
 Location/Qualifiers
 1. 759
 /organism="Mus musculus"
 /strain="C57BL/6 x CBA"
 /db_xref="taxon:10090"
 /clone="IMAGE:1278139"
 /clone_lib="Stratagene mouse lung 937302"
 /sex="female"
 /tissue_type="lung"
 /dev_stage="6-8 month old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: lung; Vector: pBluescript SK-; Site 1: ECORI ; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. 6-8 month old female lung and 1.5 year old male lung were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "

BASE COUNT 244 a 168 c 153 g 193 t 1 others
 ORIGIN
 Query Match 11.9%; Score 505.2; DB 9; Length 759;
 Best Local Similarity 85.3%; Pred. No. 3.4e-79;
 Matches 598; Conservative 0; Mismatches 99; Indels 4; Gaps 3;
 QY 2118 ctgagcagtgctgtcgaagtattatctgtccctccctaccatcatgagagattcagtgag 2177
 Db 1 CTTGAAATGGCTGTGTCAAGTTTTCTGTCCCTCCCTACCACCATTGGAGATTCAGTGGAG 60
 QY 2178 ttcaactgtacagaaaccttcacaatgattggacatgcagtagtttcttcattagtgga 2237
 Db 61 TTCACTGTGCAGAAACCTTCACAAATGATTGGACTTGGGTGAGTTCTTCCCTTAGTGGGA 120
 QY 2238 agtgagcagagcttccctcctcaatgattgttgcacagatcaactgagaagtgtaaaagcccg 2297
 Db 121 AAGTGGACCCAGCTTCTCTAAATGTTGTCAACAGACCACTGGGAAAGTGTAGAGTGTCTG 180
 QY 2298 aagtcacactggcagatgcaattcctcaataagaatgaatttaatacatttagt 2357
 Db 181 AAGTCAACTGCATAGAGAAGCAATTAACCAAAAGAAATGAATTTTCAGCAATACTCCACC 240
 QY 2358 gtgagttacagatgtagacaaaagcagagagatgaacattcaactctgcataatggaga 2417
 Db 241 ATGTATTACAAATGTAGAGACAGCAGGAGTATGAACACTCAATCTGTATCAATGGAAAA 300
 QY 2418 tgggactctgaaacaaactgtacaaagaaagattctgcctcctcccccgcagattcca 2477
 Db 301 TGGAAATCTGAACCAACTGTACAAGGAAAAATCTGCTCCCTCTCCACCACAGATTCCA 360
 QY 2478 aatgcccaagtgattgaaaccaccggtgaaatacttggatggagagaaaaagtattgtctt 2537
 Db 361 AATACCCTAGTGTATGAACCACTGTGAAATACTTGGATGGAGAAAATATATCTGTCTTT 420
 QY 2538 tgccaagatggttacctaacctcagggcccgagagaagaatggtgtgtaaacatggaagtg 2597
 Db 421 TGCCAAGACAATTAACCTAAGTACAGACCAAGAGAAATGATGTGCAAGAGATGAAGGTGG 480
 QY 2598 cagctgtaccacgctgcacgggaaaaattccatgttccacgccccctaaaaattgaacat 2657
 Db 481 CAGTCAATTACCTCACTGCATTTGAAAAAATTCATGTTCCAGCCCTTAAATAGAACAT 540
 QY 2658 ggaatctattaagtcgcccaaggtcctcagagagagagagagatttaattagtcacagct 2717
 Db 541 GGATCTATTAAATTAACCCAGAACTTCTGGA-AAAAGAGAAAAATCCATTGTGANGTCAGCAGT 599
 QY 2718 tatgaacacggaactacattcagctattgtgtgagatggatgcagatatctgaagaa 2777
 Db 600 CACGAACATGGAACTACATTCAGTATGCTGTGATGATGGGTTC-AGATATCTGAAGAA 658

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 16:09:45 ; Search time 537.93 Seconds
(without alignments)
13497.725 Million cell updates/sec

Title: US-09-316-163-1

Perfect score: 4229

Sequence: 1 tcgagtcactgtctccagc.....aaaaaaaaaaaaaaaaaaaa 4229

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	437.4	10.3	767	18	AAV02792 Human partial comp
2	432.4	10.2	525	21	AAA43501 Mouse secreted exp
3	427.2	10.1	649	18	AAV02790 Human partial comp
4	405	9.6	688	18	AAV02794 Clone PRBS3FH2910
5	376.8	8.9	581	18	AAV02791 Clone PRBS9FH410 C
6	334.4	7.9	550	18	AAV02795 Clone PRBS3FH2910
7	325.2	7.7	1293	21	AACT7947 Human cancer assoc
8	321.6	7.6	759	20	AA39793 Gastric cancer ass
9	307.2	7.3	532	18	AAV02793 Clone PRBS3FH2910

10	244.8	5.8	472	18	AAV02796 Human partial comp
11	196	4.6	2033	20	AA34737 Human complement f
12	194.2	4.6	385	18	AAV02798 Clone p253FH2576 #
13	192.6	4.6	385	18	AAV02797 Clone p253FH2576 #
14	115.6	2.7	137	22	AAF31736 Novel rat gene fra
15	113.4	2.7	2177	9	AA82402 B subunit of human
16	102.6	2.4	465	22	ABA42243 Human breast cell
17	102.6	2.4	465	22	ABA52669 Human foetal liver
18	102.6	2.4	465	22	ABA22457 Probe #923 for gen
19	102.6	2.4	465	22	AAK00926 Human brain expres
20	102.6	2.4	465	22	AAK26380 Human bone marrow
21	102.6	2.4	465	22	AAI11013 Probe #946 for gen
22	102.6	2.4	465	22	AAI32275 Probe #961 used to
23	102.6	2.4	465	22	AAI00933 Probe #924 used to
24	84.4	2.0	177	22	ABA64580 Human foetal liver
25	84.4	2.0	177	22	AAI44752 Probe #13438 used
26	80	1.9	156	22	ABA47391 Human breast cell
27	80	1.9	156	22	ABA65281 Human foetal liver
28	80	1.9	156	22	ABA32382 Probe #10848 for g
29	80	1.9	156	22	AAK13693 Human brain expres
30	80	1.9	156	22	AAK39433 Human bone marrow
31	80	1.9	156	22	AAI20246 Probe #10179 for g
32	80	1.9	156	22	AAI45448 Probe #14134 used
33	80	1.9	156	22	AAI05950 Probe #5941 used t
34	78	1.8	177	22	ABA65279 Human foetal liver
35	78	1.8	177	22	ABA32380 Probe #10846 for g
36	76.2	1.8	1929	22	AA505029 Angiotensin conver
37	74.2	1.8	335	16	AAT20816 Human gene signatu
38	71.4	1.7	177	22	ABA65280 Human foetal liver
39	71.4	1.7	177	22	ABA32381 Probe #10847 for g
40	67.2	1.6	422	22	ABA51919 Human foetal liver
41	67.2	1.6	422	22	AAI31529 Probe #215 used to
42	67.2	1.6	1194	22	AA506052 Angiotensin conver
43	66.2	1.6	494	21	AAC71094 Single nucleotide
44	66.2	1.6	494	21	AAC71100 Single nucleotide
45	66.2	1.6	494	21	AAC71115 Single nucleotide

ALIGNMENTS

RESULT 1
AAV02792
ID AAV02792 standard; DNA; 767 BP.
XX AAV02792;
XX AC
XX AC
DT 27-APR-1998 (first entry)
XX
DE Human partial complement factor H cDNA fragment 2.
KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator; SS.
XX Homo sapiens.
XX
XX WO9738136-A1.
XX
PD 16-OCT-1997.
XX
XX 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
PR 09-APR-1996; 96US-0015083.
PR 09-APR-1996; 96US-0630048.
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
XX Enfield DL, Hass GM, Kinders RJ;
PI WPI; 1997-512742/47.
DR P-PSDB; AAW39156.

CC in the exemplification of the present invention.
XX
SQ Sequence 525 BP; 150 A; 101 C; 131 G; 142 T; 1 other;

Query Match 10.2%; Score 432.4; DB 21; Length 525;
Best Local Similarity 92.5%; Pred. NO. 7.4e-93;
Matches 454; Conservative 0; Mismatches 37; Indels 0; Gaps 0

Qy	1	tcgagtcacactgctccacagatagatccaaacatgagactgtcagcaagaattatttggc	60
Db	35	tgagctcagtttggctccagaaagatccaaattatgagactgtcagcaagaattatttggc	94
Qy	61	ttaattatgagactgtttgtgtagcagaagattgtaaaggtcctctccaagagaaaatt	120
Db	95	ttaattatgagactgtttgtgtcagcagaagattgttaaaggtcctctccaagagaaaatt	154
Qy	121	cagaaattctctcaggctgtgtgtctgtaaacacattatctcagaaggaactcaggcaacct	180
Db	155	cagaaattctctcaggctgtgtgcagaaacattatccagaagggcaccacaggttacct	214
Qy	181	acaaatgcgccttgatcacccgaacacttggctactttgttaaagtatcagaagaatggag	240
Db	215	acaaatgcgccttgatcacccgaacacttggcactattgtcaaagtatcagaagaatggaa	274
Qy	241	aatgggtacctcttaacccatcaaggatatgtcgaaaaggccatgtgggcatccccggag	300
Db	275	aatgggtggctcttaacccatccaggatattgtcgaaaaggccttgtgggcatccccggag	334
Qy	301	acacacctttgggtcctttagctcggcagttgactggaattggaatttgggtgcaaaagg	360
Db	335	acacacctttgggtcctttaggctggcagttggactcgaatttgggttgggtgcaaaagg	394
Qy	361	tgtttatacatgtgatagaagggtaccacactattagtgaaaattgattaccgtggaattgtg	420
Db	395	tgtttatacctgtgatgatgggtatcaactattagggtgaaattgattaccgtggaattgtg	454
Qy	421	atgcagatgggtggaccacatgatattccaatatgtcaagtttggaaagtgtctgtccaagtga	480
Db	455	gtgcagatgggtggatcaatgataattccacatactccaactctgaanttttggaaagtgtcttga	514
Qy	481	cagaactcggag	491
Db	515	cagaactcggag	525

RESULT 3
AAV02790
ID AAV02790 standard; DNA; 649 BP.

DT 27-APR-1998 (first entry)

DE Human partial complement factor H cDNA.

KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator; ss.

OS Homo sapiens

XX PN WO9738136-A1

XX
PD 16-OCT-1997

XX
PF 09-APR-1997. 07W0-UC05710

XX
PP 06-MAR-1987. 0200C-0010401

PR 09-APR-1996; 96US-0015083.

PR 06-MAR-1997; 97US-0038614.

PA (BARD-) BARD DIAGNOSTIC SCI

XX PI XX DR DR XX PT PT PT XX PS XX CC CC CC CC CC CC CC CC XX SQ

```
Query Match      10.1%; Score 427.2; DB 18; -Length 649;
Best Local Similarity 78.7%; Pred. No. 1.4e-91;
Matches 510: Conservative 0; Mismatches 138; Indels 0; Caps 0;
```

Qy	369	acatgtgatgaaggtgaccacactactaggTgaattgattaccgtgaattgtatgcagat	428
Db	1	acatgtaatgagggttatcaattgctaggTgagattaaattaccgtgaattgtgacacagat	60
Qy	429	gggtggaccaatgatattccaatatgtgaagtTgtgaagtgttgcagtgacagaactg	488
Db	61	ggagcccaatgatactcctatattgaaagtTgtgaagtgttaccagtgacagaccac	120
Qy	489	gagaattggaagaattgtgagtTgtgagccgaaccagacaggagataattattttggacag	548
Db	121	gagaatggaaaaattgtcagtTgcaattggaaccagatcggaataccattttggacaa	180
Qy	549	gtgtgacgctttgaatgcacactccggctccaagattTgaaggacagaaaaaataatgcactgc	608
Db	181	gcagtacggtttgtatgaactcagctacaagattTgaaggagatgaagaatgcattgt	240
Qy	609	tctaataatggcctctggagcaatgaaaagccacagtTgtgtgaaatttcttgcctgcaca	668
Db	241	tcagacgagtgttttgagTaaagagaaaccaaagtTgtgtgaaatttctatgcacaaatcc	300
Qy	669	ccacgadtTgaaaattggagatgggtatatactTgaaaccagattTacaaggagagaatgaaga	728
Db	301	ccagatgttaataatggatctcctatctcagaagattTatttataaggaggaatgaacga	360
Qy	729	ttccaataataatgtaagcaaggtttgtgtacaagaagaaggagatgctgtctgcacg	788
Db	361	ttccaataataatgtaacatggtgttatgaatacagTgaaaggaggagatgctgtatgcact	420
Qy	789	ggttctggatggaaactcactcagcctctcctgtTgaagaaatgacatgtttgactccatatatt	848
Db	421	gaatctggatggcgttcgttgccttctatgtgaagaaaaatcatgtgataatccttatatt	480
Qy	849	ccaaatggatatcacacactcacaggtataaacacagaattTgatTgaaatacagatat	908
Db	481	ccaaatggTgactactcacctttaaggattaaacacagaaactggagatgaaatacgcgtac	540
Qy	909	gaatgTaaaaatggctctctatcctgTgaaccccgatcacctgtttcTaaagtTgataattact	968
Db	541	cagtTgaaaatggtttttatcctgTcaaacccggggaaatacagagccaaatgcacagatgact	600
Qy	969	ggctggatccctgcctccaagatgtagctTgaaaccccttgtgattttcca	1016
Db	601	ggctggataactgtctccgagatgatacctTgaaaccccttgtgattttcca	648

CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
CC Agents that may modulate this antigen could be used in the manufacture of
CC a medicament for the treatment of a tumour cell.
XX
SQ Sequence 581 BP: 198 A; 85 C; 139 G; 159 T; 0 other:

Query Match 8.9%; Score 376.8; DB 18; Length 581;
Best Local Similarity 78.7%; Pred. NO. 1.2e-79;
Matches 450; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

369	Qy	acatgtgatgaagggtaccacactattagtgtaaatgattaccogtgaatgtgatgcagat	428
	Db	1 acatgtaatgagggtgtacatctgtcagtgagatcaattaccogtgaatgtgacacgat	60
429	Qy	gggtggaccaatgatatccaatatgtgcaatgtgaaagttgtgaaagtccttgccagtgacagaactg	488
	Db	61 ggtggaccaatgatatccctatcatgtgaaagttgtgaaagtggttaccagtgaccagcacca	120
489	Qy	gagaatggaaagaattgtagtggtgcagcogacaccagaccaggaaattatttttggacag	548
	Db	121 gagaatggaaaaattgcagtagtgcaatggaaaccagatcgggaaataccattttggacaa	180
549	Qy	gtgtgacgctttgaatgcaactccggcttcaagatggaaggacagaaaagaataatgcactgc	608
	Db	181 cgagtacggtttgatgttaactcaggctacaagatggaaggagatgaagaaatgcattgt	240
609	Qy	tcataaatggcctctgagcaaatgaaaagccacagtggtggaaattttcttgcttgcca	668
	Db	241 tcagacgatggttttgggtgaagagaaaaccaaagtggtggaaatttcatgtcaaatcc	300
669	Qy	ccaagtggtgaaatggagatgggtatatactctgaaccagtttccaaaggagagaatgaaag	728
	Db	301 ccagatgtataaatggatcctctatatctcagaagatattttataaggagagaatgaacga	360
729	Qy	ttccaatataaatgtaagcaagggttttgtgtacaaagaaaggagggtgctgtctgcacg	788
	Db	361 tttccaatataaatgtaaactgggtttatgaatacagtgaaaggagagatgctgtatgcact	420
789	Qy	ggttctggatggaaatcctcagccttctgtggaagaatgacatggtttgactccatatatt	848
	Db	421 gaa tctggatggcgctccgttgccttctcatgtgaaagaaatcatgtgatatactctatat	480
849	Qy	ccaaatgggtatctacacacctcacaggtattaacacagaattgatgtgaaatcgagat	908
	Db	481 ccaaatgggtgactactcaaccttaaggattaaacacagaactggagatgaaatcacgtac	540
909	Qy	gaatgtaaaaaatggcttctctatcctgcaacccg	940
541	Db	cagtgtagaaatggtttttatcctgcaacccg	572

RESULT

RESOL 0
AAV02795
ID AAV02795 standard: DNA: 590 BP.

AAV02795;

DT 27-APR-1998 (first entry)

DE Clone PRBS3FH2910 #4.1 CFH related protein DNA fragment.

Complement factor H; tumour associated antigen; renal cancer; KW
urogenital cancer; medicament; modulator; ss. KW

OS Synthetic.

PN W09738136-A1.

16-OCT-1997.

AA 09-APR-1997: 97WO-US05710.

XY

PR PR PR PR XX PA XX PI XX DR DR XX PT PT PT XX PS XX CC CC CC CC CC CC CC XX SQ

```
Query Match          7.9%; Score 334.4; DB 18; Length 590;
Best Local Similarity 73.3%; Pred. No. 1.5e-69;
Matches 428; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
```

Qy 3021 ttgcccacatttgaaattgccaaaccgacagaaaaaataatcatacaggtcaqqa 3080

Db 3 ttgcctagcctttgaaaaatgcccataccctatcccatqqgagaaagaaqqatttqtataagccqqqt 62

Qy 3081 gaacaagtacatttcagatgtccacctccgtatctcaatqqatqqctctgacattgtcaca 3140

Db 63 gagccagctgacttacacttgcacacatatcacaaaatggatggagccagtaataataaca 122

Oy 3141 TGTGTTAATACGAAAGTGGATTGGACAGCCGGTATGCAAGATAATTCCCTATATGAATCCA 3200

123 TGCATTAAATAGCCAGATGGACAGGAAGGCCAACATGCCAGAGACACCTCCTATATGAATCCG 182

Oy 3207 ccacatqfccccaaatgcctactatatccacaaccacaatacattccc

182

[illegible][illegible][illegible]

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523</
--	---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-------

[illegible][illegible][illegible][illegible]

423 ccaccagcttgaggtaccgaatgccagaacttgatcaacttgagggtaaccaagcgataaca 482

QY 350I 5GTAGAAATGGAAGTGGTCTCAGCCACCAACCTGCTTACATGTCATGTCGTGATACCAGAA 3560

Db 483 t gtagaaatggacaatggtcagaaaccaccaaatagtctacatccgtgtgtaatatcccg 542

QY 3561 gatattatggaaaaaacataataatagttctcagatggaggaaaa 3604

Db 543 gaaattatggaaaaattataacatagcattaaggtggacagccaa 586

RESULT 7
ID AAC77947
XX AAC77947 standard; cDNA; 1293 BP.
AC AAC77947;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:341.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnary; immunomodulator;
KW antidiabetic; antiaesthetic; antirheumatic; antiarthritic; antiviral;
KW dermatological; antithyroid; antiallergic; antibacterial; cardiant;
KW vasotropic; antipsoriatic; thrombolytic; coagulant; nootropic;
KW immune disorder; haematopoietic cell disease; gene therapy; inflammation;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
XX
OS Homo sapiens.
XX
PN WO200055350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587533/55.
DR P-PSDB; AAB43738.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 1; Page 892; 2352pp; English.
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnary; immunomodulator;
CC antidiabetic; antiaesthetic; antirheumatic; antiarthritic;
CC antinflamatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC neotropic; vasotropic; antipsoriatic and angiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 1293 BP; 427 A; 242 C; 247 G; 367 T; 10 other;

Query Match 7.7%; Score 325.2; DB 21; Length 1293;
Best Local Similarity 67.7%; Pred. No. 3e-67;
Matches 459; Conservative 8; Mismatches 206; Indels 5; Gaps 1;

QY 3174 tgcaagataatcctctgtgtgaatccaccacatgtgccaatgtctactatactacaagg 3233
DB 512 tccactgacactctctgtgtgaatccgccacagtcacaaatgcyatatastgtcgaga 571
QY 3234 cacaagactaaatataccatctctgtgtgacaaaagtcagttatgactgttaataaaacctttgaa 3293
DB 572 cagatgagtaaatatccatctctgtgtgagagtcagttatataatgttggagcccttatgaa 631
QY 3294 ttatttgggaagtgaagtgtgccaacacgggatttgacagaaaccccgaaatgc 3353
DB 632 atgtttgggagtgaagtgatgttttaaatggaacctggagcgaacccctcaatgc 691
QY 3354 aaagattcaacagggaatgtgtgctctccacatttgacaatggagacacactcc 3413
DB 692 aaagattctacrggaatgtgtgcccctccacatttgacaatggagacacttca 751
QY 3414 ttgtcattaccagttatgacacattatcactcagttgaatcaatgccagaaactatt 3473
DB 752 ttcccgttgctagttatgctccagcttcactcagttgagtaacaaatgccagaaactgtat 811
QY 3474 ctacttaagggaataagatagtaacatgtagaaatggaaatggatgctcagccacac 3533
DB 812 caacttgagggttaacagcgaataacatgtagaaatggaaatggatgctcagccacac 871
QY 3534 tgcctacatgcattgtgtatcaccagaagatattatggaaaaacataataatgtctcaga 3593
DB 872 tgcctacatgcattgtgtatcaccagaagatattatggaaaaacataataatgtcagtaagg 931
QY 3594 tggagggaatgcaaaagatttattcccaatcagggggagaatattgaattcctgtgtaa 3653
DB 932 tggacagccaaacagagctttatttggagacagtgatgagtcagtgatgtgtgtaa 991
QY 3654 cctggatagaaaaattcagaggtacacctcgttctgcacaaagtgcattgaggggtcac 3713
DB 992 cggggatattcgttctcagcgttctcagacattgcgaacacacattgtggatgggaaa 1051
QY 3714 atcaattatcccaactgtgtataaaatcgctatacaaatatttagtaaaccttatggatga 3773
DB 1052 ctggagtagtcccaactgtgcaaaaagatagaatcaatcataaartgcacacactttattca 1111
QY 3774 gaattgacatgtatattactaataacagtttgaatttaccatttaaatattgtttagctca 3833
DB 1112 gaactttagtattaaatcagttctyaatttcattt-----ttwatgtattgtttactcc 1166
QY 3834 ttctcttaataagtata 3851
DB 1167 tttttattcactacgcaaa 1184
RESULT 8
AAC39793/C
ID AAC39793 standard; DNA; 759 BP.
XX AAC39793;
AC AAC39793;
DT 02-JUL-1999 (first entry)
DE Gastric cancer associated gene.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX Homo sapiens.
XX WO9904265-A2.
XX
PD 28-JAN-1999.
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.


```
FT CDS 78..1814
FT /*tag= a
FT /note= *the coding sequence is specifically claimed
FT for in claim 2*
XX
XX
XX PN W09918200-A1.
XX
XX PD 15-APR-1999.
XX
XX PF 02-OCT-1998; 98WO-JP04448.
XX
XX PR 06-OCT-1997; 97JP-0272837.
XX
XX PA (PROT-) PROTEGENE INC.
XX (SAGA ) SAGAMI CHEM RES CENT.
XX
XX PI Kato S, Sekine S;
XX
XX DR WPI: 1999-264019/22.
XX P-PSDB; AAY09065.
XX
XX PT Human proteins with secretory signal sequences and nucleotide
XX PT sequences, useful in control of proliferation and differentiation of
XX PT cells
XX
XX PS Claims 2, 4; Page 60-64; 71pp; English.
XX
XX CC This DNA encodes a protein having homology to human complement factor H,
XX CC which plays a role in the immunological mechanism involving the
XX CC complement reaction. The protein can also be used as an antigen for
XX CC preparing antibodies against the protein. The cDNA can be used as a probe
XX CC for gene diagnosis and the gene for gene therapy, as well as for large-
XX CC scale expression of the protein. The protein may also have immune
XX CC stimulating or suppressing activity, hematopoiesis regulating activity,
XX CC tissue growth activity, activin/inhibin activity, anti-inflammatory
XX CC activity, tumour inhibition activity, chemotactic/chemokinetic activity,
XX CC receptor/ligand activity, etc. The protein is identified by the presence
XX CC of a hydrophobic N-terminal secretory signal region, knowledge of the
XX CC protein function is not required, as in e.g. methods of expression
XX CC cloning.
XX
XX SQ Sequence 2033 BP; 682 A; 375 C; 374 G; 602 T; 0 other;

Query Match 4.6%; Score 196; DB 20; Length 2033;
Best Local Similarity 69.7%; Pred. No. 1.8e-36;
Matches 265; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 1350 ccgaatgcgtccgtatcaagaactgttcagtcagatagataagaattgaaatgggttt 1409
DB 1056 ccaacagtcctcgcacgaacatgctcaaaatcagatagataagaattgaaatggaltc 1115
QY 1410 tttctgaatcgtatatacatatgcctctataataagaacacggtatagatgtaaacag 1469
DB 1116 attctgaatcctctctatttatattttaaaagaatacaataataatgtaaacca 1175
QY 1470 gqatgttaacaaataccggaagaataatcagaaataacttgccttcacagatgatgg 1529
DB 1176 gqatgttaacacgagcaggaattctcaggttcaatgattgttcgaaatggatgg 1235
QY 1530 tcactcgaccctcagtcattaaagtcttgatgatacgtctgatttgagaattctatgact 1589
DB 1236 tcagcacaacaaattgcattaaatttgatagtcgtctgttttgagaattccagagcc 1295
QY 1590 aagaataataacacatggtttaaactaatgacaaatagacatgatagtgtcacattgga 1649
DB 1296 aagagtaatggcgtcggttttaagctccatgacacatggactgacgaatgctacatgga 1355
QY 1650 tatgaaatgaataataacacaaagcctctataacatgatactatgatgagtggtct 1709
DB 1356 tatgaaatcagttatggaacacacaggttccatagctgtggtggaagatgggtggtcc 1415
QY 1710 agtacacccctcctgttatga 1729
```

```
DB 1416 catttcccaacatgttataa 1435
| | | | | | | |
RESULT 12
AAV02798
ID AAV02798 standard; DNA; 385 BP.
XX
XX AC AAV02798;
XX
XX DT 27-APR-1998 (first entry)
XX
XX DE Clone pZS3FH2576 #1/11 CFH related protein DNA fragment.
XX
XX KW Complement factor H; tumour associated antigen; renal cancer;
XX KW urogenital cancer; medicament; modulator; ss.
XX
XX OS Synthetic.
XX
XX PN W09738136-A1.
XX
XX PD 16-OCT-1997.
XX
XX PF 09-APR-1997; 97WO-US05710.
XX
XX PR 06-MAR-1997; 97US-0812481.
XX PR 09-APR-1996; 96US-0015083.
XX PR 09-APR-1996; 96US-0630048.
XX PR 06-MAR-1997; 97US-0038614.
XX
XX PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
XX PI Enfield DL, Hass GM, Kinders RJ;
XX
XX DR WPI: 1997-512742/47.
XX P-PSDB; AAW39162.
XX
XX PT Treating or screening for cancer, e.g. renal or urogenital cancer -
XX PT by modulating or detecting tumour associated human complement Factor
XX PT H related antigen, or nucleic acid encoding it
XX
XX PS Example 6B; Fig 8A; 104pp; English.
XX
XX CC This partial cDNA sequence is present in clone pZS3FH2576 #1/11 and
XX CC encodes a complement factor H related protein with homology to a region
XX CC of the human tumour-associated complement factor H (CFH). The detection
XX CC of this protein and a CFH antigen can be used in screening or for the
XX CC treatment of renal or urogenital cancer, e.g. bladder, cervical or
XX CC prostate cancer. Agents that may modulate this antigen could be used in
XX CC the manufacture of a medicament for the treatment of a tumour cell.
XX
XX SQ Sequence 385 BP; 128 A; 68 C; 81 G; 108 T; 0 other;

Query Match 4.6%; Score 194.2; DB 18; Length 385;
Best Local Similarity 70.2%; Pred. No. 2.5e-36;
Matches 287; Conservative 0; Mismatches 98; Indels 24; Gaps 1;

QY 2526 gtatctgtctcttgcgaagatggttacctaactcaggcccgagagaaatggtgtgtaaa 2585
DB 1 gtatctgtctcttgcgaagaaattatcattcagggaaggagagaaattacatgcaaa 60
QY 2586 catggaagggtgcagtcgttaccacgctgcacggaataattccattccagcccccct 2645
DB 61 gatggaagatggcagtcataaccactctgtgtgaaataattccattctcacaaccct 120
QY 2646 aaaaatgaacatgcatctatttaagtcccgaggtcctcagaagagagagatttaatt 2705
DB 121 cagatagaacacggaacatttaattcattcctcagggtcttcacaaga----- 164
QY 2706 gagtcacagcagttatgaacacggaactacattcagcttctgtgagatgattcaag 2765
DB 165 -----aattatgcacatgggactaaattgattactctgtgaggtggttcagg 216
```

```
QY 2766 atatctgaagaaatagggtaacctgcaacatgggaatgagctctctgctgtgt 2825
DB 217 atctctgagaataatgaaacaacatgctacatgggaatggagttctccacctcagtg 276
QY 2826 gttggaataccttgggaggggaggttcaattcctctctggtattgtttctcatgaacta 2885
DB 277 gaaggctcctctgtaaatctccacctgagattctctcatggtgtgtgtagctcaatgtca 336
QY 2886 gaaatttaccatgagagaggttaccatatacaataatttctgaaggct 2934
DB 337 gacagttatcagatgggagaaggttaccgtacaaatgtttgaagggt 385

RESULT 13
AAV02797
ID AAV02797 standard; DNA; 385 BP.
XX
AC AAV02797;
XX
DT 27-APR-1998 (first entry)
XX
DE Clone pZS3FH2576 #3 CFH related protein DNA fragment.
XX
KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator; ss.
XX
OS Synthetic.
XX
PN WO9738136-A1.
XX
PD 16-OCT-1997.
XX
PF 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
PR 09-APR-1996; 96US-0015083.
PR 09-APR-1996; 96US-0630048.
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
PI Enfield DL, Hass GM, Kinders RJ;
XX
DR WPI; 1997-512742/47.
DR P-PSDB; AAW39161.
XX
PT Treating or screening for cancer, e.g. renal or urogenital cancer -
PT by modulating or detecting tumour associated human complement Factor
PT H related antigen, or nucleic acid encoding it
XX
PS Example 6B; Fig 8A; 104pp; English.
XX
CC This partial cDNA sequence is present in clone pZS3FH2576 #3 and
CC encodes a complement factor H related protein with homology to a region
CC of the human tumour-associated complement factor H (CFH). The detection
CC of this protein and a CFH antigen can be used in screening or for the
CC treatment of renal or urogenital cancer, e.g. bladder, cervical or
CC prostate cancer. Agents that may modulate this antigen could be used in
CC the manufacture of a medicament for the treatment of a tumour cell.
XX
SQ Sequence 385 BP; 127 A; 68 C; 82 G; 108 T; 0 other;

Query Match 4.6%; Score 192.6; DB 18; Length 385;
Best Local Similarity 69.9%; Pred. No. 6e-36;
Matches 286; Conservative 0; Mismatches 99; Indels 24; Gaps 1;

QY 2526 gatatctgttttcccaagatgggttacctactcaggccaggaagaatggtgtgtaa 2585
DB 1 gatatctgttttcccaagaaattatctaatccaggagggaagaataatcacatgcaa 60
QY 2586 catggagggtgcagctgtaccagctgcacggaagaaattccatgttcccagccccct 2645
```

```
DB 61 gatgaagatggcagtcacatccactctgtgtgaaaaaattccatgttcacaccacct 120
QY 2646 aaaaatgaacatggatctatttaagtcgcccagggtctctcagaagagagagatttaatt 2705
DB 121 cagatagaacacggaaccatttaattcatccagggtcttcacaaga----- 164
QY 2706 gagtccagcagttatgaacacggaactcattcagctattgctgtagatgagattcaag 2765
DB 165 -----aattatgcacatgggactaaattgagttatactgtgaggggtgttcagg 216
QY 2766 atatctgaagaaataggglaacctgcaacatgggaatgggaatgagctctctgctgtgt 2825
DB 217 atatctgaagaaatgaaacaacatgctacatgggaatggagttctccacctcagtg 276
QY 2826 gttggaataccttgggaggggaggttaccatatacaataattggtattgtttctcatgaacta 2885
DB 277 gaaggctcctctgtaaatctccacctgagattctctcatggtgtgtgtagctcaatgtca 336
QY 2886 gaaatttaccatgagagaggttaccatatacaataattggttctgaaggct 2934
DB 337 gacagttatcagatgggagaaggttaccgtacaaatgtttgaagggt 385

RESULT 14
AAF31736/c
ID AAF31736 standard; DNA; 137 BP.
XX
AC AAF31736;
XX
DT 10-APR-2001 (first entry)
XX
DE Novel rat gene fragment ACETA48.
XX
KW Rat; ACETA; acetaminophen; differential gene expression;
KW hepatotoxic agent; pericentral hepatic necrosis;
KW non-steroidal anti-inflammatory drug; ds.
XX
OS Rattus sp.
XX
PN WO200102609-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US40292.
XX
PR 02-JUL-1999; 99US-0142335.
PR 29-JUN-2000; 2000US-0607539.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Gould-Rothberg BE, Dipippo VA, Milroy L, Daniels KK;
XX
DR WPI; 2001-138157/14.
XX
PT Screening a test agent for hepatotoxicity, useful for identifying
PT agents that damage the liver, comprises measuring differential
PT expression of one or more ACETA nucleic acids in a test cell population
XX
PS Disclosure; Page 21; 68pp; English.
XX
CC The present sequence is an ACETA nucleic acid whose expression is
CC modulated at least 8-fold in the presence of acetaminophen. The
CC differentially expressed ACETA nucleic acids may be used to identify
CC agents that damage the liver, i.e. hepatotoxic agents that cause, for
CC example, pericentral hepatic necrosis, and to identify poisoning
CC associated with ingestion of non-steroidal anti-inflammatory drugs,
CC such as acetaminophen. These nucleic acids may also be used as
CC hybridisation probes to identify ACETA-encoding nucleic acids and
CC fragments for use as polymerase chain reaction (PCR) primers for
CC amplification or mutation of ACETA nucleic acids. ACETA proteins may
CC be used as immunogens to raise anti-ACETA antibodies.
```

```

XX SQ Sequence 137 BP; 32 A; 29 C; 33 G; 43 T; 0 other;
Query Match 2.7%; Score 115.6; DB 22; Length 137;
Best Local Similarity 96.7%; Pred. No. 7.9e-18;
Matches 118; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2445 aaaaatttgcctctcccccgcagagattccaaatgcccagtgattgaaaccaccgtg 2504
DB 122 AAAAATTTCGCCCTCTCCCCACAGATTCCAAATGCCCAAGTGATTGAAACACAGTG 63
QY 2505 aaatcttgatggagaaaaatctctgttcttgcgaagatggttacctaactcaggcc 2564
DB 62 AAATACTTGGATGGAGAGAGATCTGTTCTTGTGCCAAGATGGTTACCTAACTCAGGCG 3
QY 2565 cc 2566
DB 2 CC 1

RESULT 15
ID AAN82402 standard; DNA; 2177 BP.
XX AC AAN82402;
XX DT 26-NOV-1990 (first entry)
XX DE B subunit of human Factor XIII.
XX KW B sub-unit; factor XIII; blood clot stabilisation; fibrin polymer;
XX KW crosslink; scleroderma; haemorrhage; ulcerative colitis; ss.
XX OS Homo sapiens.
XX FH Key
XX FT CDS 2..1521 Location/Qualifiers
XX FT /*tag= a
XX FT /product=b subunit
XX FT misc_RNA 2..58
XX FT /*tag= b
XX FT /label=leader_sequence 59..1521
XX FT CDS
XX FT /*tag= c
XX FT /product=mature b subunit
XX PN A08778694-A.
XX PD 31-MAR-1988.
XX PF 21-SEP-1987; 87AU-0078694.
XX PR 19-SEP-1986; 86US-0909512.
XX PA (ZYMO-) ZYMOGENETICA INC.
XX PI Davie E, Seale RL, Ichinose A, Holly JA, Parker GE;
XX WPI; 1988-140637/21.
XX DR P-PSDB; AAP82921.
XX FT New DNA sequences encoding sub-units of factor 13 - and corresponding
XX FT expression vectors and transformed host cells.
XX PS Disclosure; ; p; English.
XX CC The carboxyl-terminal Thr (nucleotides 1979-1981) is followed by
XX CC a stop codon (TAG), a 187 bp noncoding sequence, and a poly(A) tail
XX CC of 9 bp. The polyadenylation or processing signal of AAAAA
XX CC was identified 19 nucleotides upstream from the poly(A) tail.
XX CC When cultured the host cells will produce the polypeptide which
XX CC can be assembled to factor 13, which stabilises blood clots by

```

```

CC crosslinking fibrin polymers. Factor 13 is useful therapeutically, e.g.
CC in cases of scleroderma, haemorrhage, ulcerative colitis etc., and
CC can be prep. in large amts. without risk of viral contamination.
CC See also AAN82401 and AAN82403.
XX SQ Sequence 2177 BP; 758 A; 358 C; 434 G; 627 T; 0 other;
Query Match 2.7%; Score 113.4; DB 9; Length 2177;
Best Local Similarity 48.8%; Pred. No. 7.7e-17;
Matches 436; Conservative 0; Mismatches 436; Indels 21; Gaps 4;
QY 1401 aatgggttttttctgaattctgattatcatatgctctctaaatagaaaaaacacggtaga 1460
DB 290 aatgggttcatctctgattgataaagtattgtataaaattcaagagaacatgcattatggt 349
QY 1461 tgtaaacaggatattgataaaataccgagaaataatcaggaaataattacttcttctcaa 1520
DB 350 tgcgttcagggtacaaaaccctggagggaaggatgaagagtgcttcaatgctctctct 409
QY 1521 gatggatggtcacctcgaccctcatgatttaagtccttgatag-----cctgta 1571
DB 410 gatggatggtctctctcaaccacctgtagaaaagaacatgaaacgtgtttggctctctgaa 469
QY 1572 ttggagaattctgactaagaataatacacatggttttaaacctcaatgacaaattagac 1631
DB 470 ttataatggaaattatttccacacacagaaaacatttcaagtgaaggaacaaagtacaa 529
QY 1632 tatgaatgcacatggatgataaaatgaataataacataacacaaaggctctataacatgt 1691
DB 530 tacgaatgctcactggtcactacacagctggagggaagaacagagaggtagaatgt 589
QY 1692 acttatgatggatggtctagtagcacccctctgttatgaaagagaatgcagcatctccctg 1751
DB 590 ctacatacggatggtctctcacacaaaatgtaccaaaattaaagtctctctttaaga 649
QY 1752 ttacaccaagattagttgttttccacagagaagtaaaatacaaaattggagattcgtgt 1811
DB 650 ttaattgaaaatggttattttctctctgtaagcaaacctatgaagaaaggagatgctgt 709
QY 1812 agttctctctgcgttca--ggacacagagttggagcagatttaagccaagctaccac 1868
DB 710 cagttttctgcatgaaaattattattctcaatgctgatttcaatcagtgctataac 769
QY 1869 ttggatggtcccttaatttcccaacgtgtgaaggccaagtataaaatcatgtgaccaacct 1928
DB 770 ttgtgttggtaccagaatctctctgtatgcgaaggagaagaacacagatgctctctcca 829
QY 1929 cttgaaatcccgaaatggggaataaaaggggaacaaaaaaagttagaacacagccatgtgac 1988
DB 830 cctctgcccataaaactccaaaattca---aacacattcaacaacttatcgtcatggagaa 886
QY 1989 gtggtggaataatgattgcgaacacttagatttctactgaagggaacccaataaaatccagtt 2048
DB 887 atagttcatatagaattggaacttaattttgagatccatgggtcagcagaataacgttgt 946
QY 2049 gttgacgggaagtgacaaaggttgcgatatcgcttg-----agtatgagagaacatgt 2102
DB 947 gaagtggaataatggacagaacctccaaaatgcattgaaggacagagaagtagcctgt 1006
QY 2103 ggagacctctctgaacttgagcatggtctgtctgaagttatctgtccctccctaccatcat 2162
DB 1007 gaggaaaccacctcttattgaaaatggtgcagcaaatattacactctaagatttatacaat 1066
QY 2163 ggagattcagtggaattcacttctgacagaaaccttcacaatgattggacatgcagtagtt 2222
DB 1067 ggggataaaggacatatgcatgtaaaaggcggtccctctccatggatcccaatgagata 1126
QY 2223 ttctgcattagtggaagtggaacccgagcttccctcaatgtgttgcaacagatca 2275
DB 1127 acttgtaactgtggaaaatggacacttctcctgagtggtgttgaaataatga 1179

```

Search completed: August 30, 2002, 19:15:26
Job time: 11141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 19:06:49 ; Search time 6225.59 seconds
(without alignments)
2910.950 Million cell updates/sec

Title: US-09-316-163-2
Perfect score: 866
Sequence: 1 tcagatcaactgtcccca.....ttgaattacattgaaaaa 866

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg90_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

A91884
LOCUS A91884
DEFINITION Sequence 2 from Patent WO9823638.
ACCESSION A91884
VERSION A91884.1 GI:6740761
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.

REFERENCE
AUTHORS Schwaebler, W. and Sim, R. B.
TITLE COMPLEMENT INHIBITOR

JOURNAL
Patent: WO 9823638-A 2 04-JUN-1998;
Schwaebler Wilhelm (GB); Univ Leicester (GB)

FEATURES
source

Location/Qualifiers
1..866
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 273 a 154 c 204 g 235 t
ORIGIN

linear PAT 22-JAN-2000

A91884 Sequence 2
A91883 Sequence 1
AJ320522 Rattus no
A91885 Sequence 3
M12660 Mouse CFh 1
X07523 Human mRNA
Y00716 Human mRNA
A91886 Sequence 4
X04697 Human mRNA
BC012610 Homo sapi
X98697 B. taurus nr
AC096417 Rattus no
AL049744 Human DNA
AL161735 Homo sapi
M29010 Mouse compl
M29009 Mouse compl
AF348144 Rattus no
AF348147 Rattus no
AF348148 Rattus no
J02891 Mouse compl
M29008 Mouse compl
L21703 Paralabrax
M31979 Mouse facto
I66494 Sequence 14
U56979 Human compl
AC097078 Rattus no
AX050034 Sequence
AX350821 Sequence
AL158158 Human DNA
AX345792 Sequence
AC096300 Rattus no
M31980 Mouse facto
AF429315 Homo sapi
BC003411 Homo sapi
AF212235 Homo sapi
BC008198 Homo sapi
AC022107 Homo sapi
AC010273 Homo sapi
AC023629 Homo sapi
AC024583 Homo sapi
AL079279 Homo sapi
AC087887 Homo sapi
AC024319 Homo sapi
AC079597 Homo sapi
AC092267 Homo sapi

ACCESSION AJ320522
VERSION AJ320522.1 GI:15485712
KEYWORDS complement inhibitory factor H, FH gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 4256)
Demberg, T., Goetze, O., and Schlaf, G.
Rat complement factor H: molecular cloning, sequencing and
expression in tissues and isolated cells
unpublished
2 (bases 1 to 4256)
Demberg, T.
Direct Submission
Submitted (07-AUG-2001) Demberg T., Immunology, Georg August
University of Goettingen, Kreuzbergstr 57, 37073 Goettingen,
GERMANY

FEATURES
source
1..4256
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="liver"
/clone_lib="Lambda Triplex"
/dev_stage="ages 10 - 12 weeks"
/tissue_lib="CLONTECH rat liver 5' stretch plus cDNA
library"
24..4209
/gene="FH"
24..3734
/gene="FH"
/codon_start=1
/product="complement inhibitory factor H"
/protein_id="CAC67513.1"
/db_xref="GI:15485713"
/translation="MRLSARIILWLTVCVAEDCKGPPPRENSEILLGSWSSEQLYSE
GTQATYKCRPGYRLTGLIVKKNGEWVPSNPSRIKRCPCGHPGDTFGSFLAVGS
EFEGAKVYVQYQQLLGEIDYRECDADGWINDIPICEVVKCLPVLLENGRIVSG
AAEPDQYVYQVVFECNGKIEGQKEMHCSNGLWSNPKOCVETSLPLPRVNG
DGIYLPKVYKNERFOYKQGVYKRGDAVCTGSGWNPQSCSEEMTCLTPYIPNGI
YTPHRIKHRIIDEIYRECKNGLYPATRSPVSKTITCWIAPRSLKPCDFPQPKHGR
LYEESRPYFPVIGREYSYCDNGFTTPSQSDWYLRKTVNGWEVPCRLQCFIH
SVIEGSIYWRRTIEGSAKVQCHSYLSPNGQDTILCTENGWSPPKVRIRKTSV
CDMPVENAMTKNNTWFKLNDKLDYECHEIGYENYKHTGSICTYDWSSTPSCYE
RECSIPLLHQDLVFPREVYKVGDSLSFSCSRGHRGADLVQCYHFGWSNFPPTCEG
QVKSODPLEIPNGEIKTKKVEYSHGDVVEYDCKPRFLKGNKIQCYDQKWTLLPI
CVEYERTCDLPELHSGVSKLSPYPYHHGDSVEFTCTETFMIGHAVVFCISGRWTEL
PQCVAITQLEKCAKPSGTGIDAIHPNKNFNNFSVSRCKRQKOEYEHISICINGRWEL
EPNCTRNEKRCPPPPQIPNAQVIETTVKYLDEKVSVLCDQGYLTQGPEDMCKHGR
WQSLPTEKTEKIPCSOPKIEHGSIKSPRSEERDLIESSEYHGTFTSYVCDGDFRIS
EENRTCMKMGWSLPCRCVGPCCGPPSIPGLIVSHELESYQVGEVYNGSEGGID
GPATIKCVGGOWSPPKCIKTDCLNLTPEAKTEKKKSYRSGEQTVPKPPPYRM
DGSIVTCVNTKWIQVPCVCKDSCNVPHPVNAIILRHKTKYPSGQVRYDCNKPFE
LFWGSGDVPTGTWTEPKCKDSTGKCPPPPIDNGDITSLSLPVTVAPLSVSYQCC
NYLLKGNKIVTCRNGKWSOPPTCLHACVIPEDIMEXHNIVLRWENAKIYSQSGENI
EFMKPGYRFRGSPFPRTKICIEGHIYPTCV"
4194..4199
/gene="FH"
4209
/gene="FH"
polyA_signal
polyA_site
BASE COUNT 1423 a 788 c 892 g 1153 t
ORIGIN

Query Match 78.4%; Score 678.8; DB 10; Length 4256;
Best Local Similarity 99.7%; Pred. No. 3.7e-168;
Matches 680; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 10 ctgctccagatagatcccaagacatgagactgctcagcaagaattatttggtattattat 69

|||||
Db 1 CTGCTCCAGATAGATCCAGACATGAGACTGTGACGAGAATATTATTTGCTTATATTAT 60
Qy 70 ggaactttgttagcagaagattgtaaaggctcctcctcaagagaaaatttcagaaattc 129
Db 61 GGACTCTTTTGTGTAGCAGAGATTGTAAAGGCTCTCTCCAGAGAAAATTCAGAAATTC 120
Qy 130 tctcaggttcgtgctgaacaactatatcagaaggcactcaggccaacctacaaatgcc 189
Db 121 TCTCAGGTTGGTGTGAACAACTATATTAGAAGGCACTCAGGCAACCTACAAATGCC 180
Qy 190 gccctggatccagaaacacttggctactattgtaaaaatgatagaagaatgagaaatgggtac 249
Db 181 GCCCTGGATACCGAACACTTGGTACTATTGTTAAAGTATGCAAGAAATGGAGAATGGGTAC 240
Qy 250 cttctaaccatcaagatatgtcgaaagccatgtgggcatccgagagacacacct 309
Db 241 CTTCTAACCCATCAAGGATATGTGGAAAGGCCATGTGGCATCCCGGAGACACACCT 300
Qy 310 ttgggtcccttttaggtggcagttggatctgaatttgaatttggcgaagggtgtttata 369
Db 301 TTGGGTCTCTTAGGCTGGCAGTTGGATCTGAATTTGATTTGGTGCAGAGTTGTTTATA 360
Qy 370 catgtgatgaagggtaccactatttagtgaaattgattaccgtgaattgtgacgagatg 429
Db 361 CATGTGATGAAGGGTACCAACTCTTAGGTGAAATTTGATTACCGTGAATGTGATGCAGATG 420
Qy 430 gdtgaccaatcatattccaatatgtgaattgtgaagtgtgagtgcttgcagtgacagaactgg 489
Db 421 GGTGGACCAATGATATTCGAATATGTGAAGTGTGTGAAGTGTGTGCAGTGACAGAACTGG 480
Qy 490 agaattgaagaattgtgagtggtgcagccgaaaccagaccaggaataattatttggscagg 549
Db 481 ACAATGGAAGAATTGTGAGTGTGTGCAGCGCAACACAGACAGCAAGGAATATTATTGGACAGG 540
Qy 550 tggtagctttgaatgacactcgcgcttcaagattgaagagacagaagaatgacatgct 609
Db 541 TGGTAGCGTTTGAATGCAACTCCGGCTTCAAGATTGAAGGACAGAGAAGAAATGCACGTCT 600
Qy 610 cataaatggcctctggagcaatgaaagccacagctgtgtgaaatttcttgcctgcccac 669
Db 601 CAGAAAATGGCCTCTGGACCAATGAAGCCACAGTGCTGTGGAAATTTCTTGTCTCTCCAC 660
Qy 670 cagagttgaagaatggagatgg 691
Db 661 CACGAGTTGAAAATGGAGATGG 682

RESULT 4
A91885
LOCUS
DEFINITION Sequence 3 from Patent WO9823638.
ACCESSION A91885
VERSION A91885.1 GI:6740762
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 2715)
AUTHORS Schwaebler, W. and Sim, R. B.
TITLE COMPLEMENT INHIBITOR
JOURNAL Patent: WO 9823638-A 3 04-JUN-1998;
FEATURES SCHWAEBLE WILHELM (GB); UNIV LEICESTER (GB)
source Location/Qualifiers
1..2715
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 863 a 518 c 618 g 716 t
ORIGIN

Query Match 76.1%; Score 659; DB 6; Length 2715;
Best Local Similarity 94.5%; Pred. No. 6.1e-163;

Matches 683; Conservative 0; Mismatches 40; Indels 0; Gaps 0;			
QY	1	tcagtgcaactgctccagatagatccaagacatgagactgtcagcaagaattatttggc	60
Db	1	TCAGTCAACTGCTCCAGATAGATCCAAGACATGAGACTGTGAGCAAGAAATTTATTGGC	60
QY	61	ttatatggagactgttctgtagcagaagattgtaaaagtcctcccaagagaaaatt	120
Db	61	TTATATTATGAGACTGTTGTGTAGCAGAAGATTGTAAGGTCCTCTCCCAAGAGAAAATT	120
QY	121	cagaaattctcaggttcggtctgaacaaactataticcagaagcactcaggaacatt	180
Db	121	CAGAAATTCCTCAGGTTCGTGTCTGTAACAATAATATTAGAAGCCTCAGGCAACCT	180
QY	181	acaaatgcgcgccttgataccgaacacttggtactattgtlaaaagtatgcaagaatggag	240
Db	181	ACAAATGCCGCCCTGATACCGAACACTTGGTACTATTGTAAAAGATATGCAAGATGGAG	240
QY	241	aatgggtactcttaacccatcaagatatgtcggaaggccatgtggcatcccgagg	300
Db	241	AATGGGTACTCTTAACCCATCAAGGATATGTCGGAAGAGGCCATGTGGGCATCCCGGAG	300
QY	301	acacaccccttgggtcctttaggctggcagtgatgtgaaattgattaccgtgaaatgtg	360
Db	301	ACACACCCCTTGGGTCCTTTAGGCTGGCAGTTGGATCTCAATTTGAATTTGGTGCAGAGG	360
QY	361	ttgtttatcatgtgatgaagggtaccacactataggatgaaattgattaccgtgaaatgtg	420
Db	361	TTGTTTATACATGTGATGAAGGCTACCAACTATTAGTCAATTTGATACCGTGAATGTG	420
QY	421	atgcagatgggtggaccatgatattccaatatgtgaagtgtgaagtgttgcagtgta	480
Db	421	ATGCAGATGGGTGGACCAATGATATTCCAATATGTGAAGTTGTGAAGTGTGCCAGTGA	480
QY	481	cagaactgagaatgaagaattgagtggtgacccaaacagaccagaccgaattatt	540
Db	481	CAGAACTGAGAATGAGAANAATTGTGAGTGTGCACCCGAACCAAGCCAGGAGGATATT	540
QY	541	tggacaggtggttacgctttgaatgcaactccggtctcaagattgaggacagaaagaa	600
Db	541	TTGACAGGTGTTACGCTTTGAATGCAACTCCCGCTTCAAGATTGAAGACAGAAAGAA	600
QY	601	tgcactgtcataaaatggcctctgagacaatgaagaagccacagtgtgtggaattctt	660
Db	601	TGCACGTGCTATAAAATGGCTCTGGAGCAATGAAAGCCACAGTGTGTGTGAAACCTT	660
QY	661	gctgcacacagcagtgtaaatggagatggatatagaaaattcagagatcacctccct	720
Db	661	GTGATTTCCACAAATCAACATGGAGCTCTGTATTATGAAGAAACCCGAGACCCCTACT	720
QY	721	ttc 723	
Db	721	TCC 723	
RESULT	5		
MUSH			
LOCUS			
DEFINITION	MUSH	4300 bp	mRNA
ACCESSION	Mouse Cfh locus, complement protein H gene, complete cds, clones		
VERSION	M12660		
KEYWORDS	M12660.1	GI:193724	
SOURCE	complement protein H; protein H; serum glycoprotein.		
ORGANISM	Mus musculus (strain C57/B10.WR) male liver cDNA to mRNA; Mus musculus (strain C57/B10.WR) male liver DNA; and Mus musculus (strain C57/B10.WR) male liver cDNA to mRNA.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 4300; 1 to 3425; 3474 to 4300)		
TITLE	Krisjensen, T. and Tack, B.F.		
	Murine protein H is comprised of 20 repeating units, 61 amino acids in length		

JOURNAL MEDLINE COMMENT	PROC. NATL. ACAD. SCI. U.S.A. 83, 3963-3967 (1986) 86233353	
Draft entry and clean copy sequence for [1], [Unpublished (1986) Scripps Clinic and Res Found, La Jolla, CA 92037] kindly provided by		
T.Kristensen, 28-JUL-1986.		
FEATURES	Location/Qualifiers	
source	1. .4300	
	/organism="Mus musculus"	
	/strain="C57/B10.WR"	
	/db_xref="taxon:10090"	
	/sex="male"	
	/tissue_type="liver"	
exon	<1..3425	
	/note="H protein"	
sig_peptide	101..154	
	/note="H protein signal peptide"	
CDS	join(101..3425,3474..3853)	
	/note="precursor"	
	/codon_start=1	
	/product="complement protein H"	
	/protein_id="AAA37759.1"	
	/db_xref="GI:387181"	
	/translation="MRLSARIILWLTVCAAECDKPPPPRENSSEILSGWSQLYPE GTQATYKCRPGVYTLGTIVKCKNGKWASNPSCRIRCKKCPGHPGDTFGSFLAVGS QFEFGAKVYTCDDGYQLGEIDYRECCAGDWINDIPLCEVVKCLPVTLENGRIVSG AAETQOEYFGQVYFECNSGFKIEGHEIKHSENGLSNEXPRCVIELCTPPRVENG DDINVKPVYKENERHYKCKHGYKPERGDAVCTGSGWSQPFCEEKRSPPYILNGI YTPHRIIHRSDDEIRYENGYEYPTVTSKCTPTGWIPIVPRCTLKPCEFPQFKYGR LYEESLRNPFVSNKYSYKDNFGSPSPGSDYDYLCTAGQWEPVCKVCVFH VYENGDSAYWEKVYVQGSGLKQCYNGYSLONGQDTMTCTENGWSPPKCI RIKTCSA SDIHIDNGLFLESSIYALNRETSYRCKQGYVTNCEISGSLTCLONGWSPOPSCLKS CDMPVPENSITKNTTWFKLNDKLDYECLVGFENYKHTKGSITCTTYGWSDTSPCYE RCSVPTLDRKLIVSPRKEGYRVGDLLEFVCHSGHRVGDPSVOCYHFGWSPPPTCKG QVASCAPPLEILNGEINGAKVEYSHGEVYKDYDKPRFLKGNPKIQCVDGNNMTLPV CIEERTCGDIPLEHSGSAKSVPPYHHGDSVEFICEENFTMIGHGSVSCISGKWTQL PKCATVDOLKCRVLKSTGIEAIKPKLTFEFTNSTMDYKCRDQYERISGICINGKWD BPNCYSKTSCTPPPOI PNTQVETTVKYLDEGKLSVLCDQNYLTQDSEMVCKDGRMQ SLPRCLIEKIPCSQPTIEHGSINLPRSSERRDSIESSEHNGTTFSYVCDGDFRIPE ENRITCYMGKWSPTPRCVLPGCPPPSPISLGTVSLELESYQHGEVYTHCSTGFGIDG PAFIICEGGKNSDPKCIKCDVLTVPKNAIIRGKSKSYRTGEQVTPRCOSPYOMN GSDTITVCNSRWIGQVCKDNDSVDPHPVNPATIVTRKNKYLHGDVRYECNKPTEL FQGVEMCMENGITWKEPKRCDTGKGGPPPIDNGDITSLSLPVYELPSVVEQCKY YLLKKKKTITCTNGKWSBPPTCLHACVIPENIMESHNIILKWRHTEKIYSHSGEDIEF CKKYGYKAROSPPTKCTINGTINPTCV"	
mat_peptide	Join(155..3425,3474..3850)	
	/note="H protein"	
intron	/product="complement protein H"	
	3426..3473	
	/note="H cds intron"	
exon	3474..4300	
	/number=2	
BASE COUNT	1401 a 814 c 902 g 1183 t	
ORIGIN	554 bp upstream of XhoI site; chromosome 1.	
Query Match	68.9%; Score 596.6; DB 10; Length 4300;	
Best Local Similarity	91.5%; Pred. No. 1.7e-146;	
Matches 632; Conservative	0; Mismatches 59; Indels 0; Gaps 0;	
QY	1	tcagtgcaactgctccagatagatccaagacatgagactgtcagcaagaattatttggc 60
Db	69	TGAGTCACTTGGTCCCAAGATCCAATATTATGAGACTGTGAGCAAGAAATTTATTGGC 128
QY	61	ttatatggagactgttctgtagcagaagattgtaaaagtcctcccaagagaaaatt 120
Db	129	TTATATTATGAGACTGTTGTGTCAGCAGAGAAGATTGTAAGGTCCTCTCCCAAGAGAAAATT 188
QY	121	cagaaattctcaggttcggtctgaacaaactatattcagaagcactcaggaacatt 180
Db	189	CAGAAATTCCTCAGGCTCGTGGTCAGAAACAATATATCCAGAGGCACCCAGGCTACT 248

Db 646 TTGTTTCAGACGATGTTTGGAGTAAGAGAAACCAAGTGTGTGCAAAATTTTCATGCCAA 705
QY 665 gccaccacgagtgtaaaatgga 686
Db 706 ATCCCCAGATGTTATAAATGGA 727

RESULT 7
HSH LOCUS 3926 bp mRNA linear PRI 08-APR-1997
DEFINITION Human mRNA for complement factor H.
ACCESSION Y00716 M32093 X07525
VERSION Y00716.1 GI:31964
KEYWORDS complement factor H; glycoprotein; regulatory protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Day/A.J.
Direct Submission
Submitted (25-FEB-1988) Day A.J., MRC Immunochimistry Unit, Dept.
of Biochemistry, University of Oxford, South Parks Road, Oxford,
OX1 3QU
2 (bases 1 to 3926)
Ripoche, J., Day, A.J., Harris, T.J. and Sim, R.B.
The complete amino acid sequence of human complement factor H
JOURNAL Biochem. J. 249 (2), 593-602 (1988)
MEDLINE 88134059
COMMENT see Immunogenetics 27:211-214(1988) concerning TYR <-> HIS at AA
384
see X07523 for truncated (1.8 kb) mRNA sequence.
FEATURES
Location/Qualifiers
1..3926
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Bg-38-1, R2a and PE3"
/tissue_type="liver"
/clone_lib="human liver cdna in pat153/PvuII/8"
74..127
74..3769
/codon_start=1
/product="factor H"
/protein_id="CAA68704.1"
/db_xref="GI:31965"
/db_xref="SWISS-PROT:P08603"
/translation="MRLAKIICLMLAICVAEDCNELPPRRNTEILFGNSDQTYPE
GTQAIYKCRPYRSLGNVIMYCRKGWALNPLRCKQKPCGHPGDTPRGFTLTGNN
VFEGVKAVYTCNEGYQLLGINRECDTNGTNDIPICEVVKCLPVTAPENGKIYSS
AMEPRDVFHGAQVFFVNCNGYKIEGDEMHCSDGFWSEKPKCVCESKSPDVING
YSLPKIKRTGDEITYQCRNGFYFATRGNTAKTSTGWIPAPRCTLKPCDYPDIKHGG
LYHENMRPFPFVAVGKYYSYCDHEFEFTPSGYNDHIHCTQDQSWPAVPCLRKCYFP
YLENGYNHGRKQVQKSIDVACHPGYALPKAQTVTTCMENGWSPTPRCIRVKTCSK
SSIDIENGFISESQVYTKALKEKAKYQCKLYGVYADGETSGSIRGCKDQWSAQPTCKIS
CDIPVEMNARTNDFTWPKLNDLTDYEDCHQYENSTGSGTSISVCGYNGWSDLPICYE
RECELPKIDVHLNPKRDKQYKVEVLKFSCKPFTIVGPNVQCIHFLGSLPDLPIK
EQVQSCGPPPELLNGVNEKTKETGEYHSEVVEYCYNPRFLMKPNKIOCVDEGWTLLP
CIVVEESTGDPIDLEHGAQLSSPPYYIGDSVFNCSESTFMIGHSITICIGWVTO
LPQVADLKLKSSNLNILEHLKNNKFEHDNSNIRYRCRGEGWHTVTCINGRWD
PEVNSAQIOLCPPPPQIPNSHNMTTLNRYDGEKVSVLCOENYLLQEGREINCKDG
RWQSIPLCKVEIKPCSQPOPIEHGTINSRSQESYAHGKLSYTCGEGFRISENETT
CYMKWSPQCEGLPCKSPPEISHGVVAHMSDQYEGEEVYKCFEGFIDGPAIAK
CLGKWSHPSPCTICDLSLPSFENAIIPMGCKDQYKAGEQVYTCATYKMDGASNY
TCINSRWTRPTCDTSCVNPPTQONAYIVSRQMSKYPSEGRVYQCRSPYEMFGDSN
VCLNGWNTPEPCKDGTGCKGPPPDINGDITSPLSVYAPASSVEVQCNLYQLEG
NKRTICRNGOWSEPCKLHPCIVISREIMENYINLALRWTAQKLYSRITGESVEFYCKRG
YRLSSRSHTLRLTCDWCKLEYPTCAK"

mat_peptide 128..3766
variation 994
/product="factor H"
/note="polymorphism"

variation /replace="c"
1277
/note="polymorphism"
/replace="t"
1409
misc_feature /note="divergence point of full length (4.3 kb) and truncated (1.8 kb) mRNA species"
2881
variation /note="polymorphism"
/replace="g"
855 g 1076 t

BASE COUNT 1306 a 689 c 855 g 1076 t
ORIGIN

Query Match 51.6%; Score 446.8; DB 9; Length 3926;
Best Local Similarity 78.4%; Pred. No. 4.6e-107;
Matches 535; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 5 gtaactgtctccagatagatccaagacatgacagctgtcagcaagaattatttggcttat 64
Db 46 GTAAATGTCTCTTAAAGATCCAAAAATGACACTTCTAGCAAGATTTATTTCCTTAT 105
QY 65 attatgactgttctgttagcagaagattgttaaggctctctccccaagagaatcaga 124
Db 106 GTATGGGCTATTGTGTAGCAGAGATTGCAATGAACCTTCTCCAAGAGAAATACAGA 165
QY 125 aattctctcaggttctgtctgaacaactatattcagaagcactcagcgaacacclacaa 184
Db 166 AATTCTGACAGGTTCTCTGCTGACCAACATATCCAGAAAGGCCACCCAGGCTATCTATAA 225
QY 185 atgccccttgatataccgaacacttggtaactattgttaaaagtatgcaagaatggagaatg 244
Db 226 ATGCCGCCCTGGATATAGATCTCTTGGAAATGTAATGTTATGCAAGGAGGAGAAATG 285
QY 245 gttacctcttaaccatcaagatatgtcgaaaaggccatgtggcgcaccccgagacac 304
Db 286 GGTGCTCTTAATCCATTAAAGAAATGTGAGAAAGGCCCTGTGGACATCTCTGGAGATAC 345
QY 305 acccttgggtcctttagctggcagttggtctgattggaatttgggtgcaaggtgtg 364
Db 346 TCCTTTTGGTACTTTTACCTTTACAGGAGAAATGTTGTTGAATATGTTGTAAGAACTGT 405
QY 365 ttatacatgtatgaagggtacaaactattagtggaattgattaccgtgaattgtatgc 424
Db 406 GTATACATGTAATGAGGGGTATCAATGCTAGTGAGATTAAATTACCGTGAATGTGACAC 465
QY 425 agatgggtgacaaatgatattcccaatatgtgaagttggaagtgcttgcacatgcaga 484
Db 466 AGATGATGGACCAATGATATTCCTATATGTGAAGTTGTGAAGTGTGTTACCGATGACAGC 525
QY 485 actggagaatggaagaattgtgagtggtgcagccgacaccagaccaggaataattatttg 544
Db 526 ACCAGAGAAATGAAAAATTTGTGAGTAGTCAATGGAACCCAGATCGGGAATACCATTTGG 585
QY 545 acaggtgttacgtcttgaatgaactccgcttcgaagtgaagtcagaagaagaatgca 604
Db 586 ACAAGCAGTACCGTTGTGTATGTAACCTCAGGCTACAGAGATTGAAGGAGATGAAGAAATGCA 645
QY 605 ctgctcataaaatggcctctgagcaatgaaagccacagattgtggaatttcttgcct 664
Db 646 TTGTTTCAGACGATGTTTGGAGTAAGAGAAACCAAGTGTGTGGAATTTTCATGCCAA 705
QY 665 gccaccacgagtgtaaaatgga 686
Db 706 ATCCCCAGATGTTATAAATGGA 727

RESULT 8
LOCUS A91886
DEFINITION Sequence 4 from Patent WO9823638.
ACCESSION A91886
VERSION A91886.1 GI:6740763

KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1532)
AUTHORS Schwabe,W. and Sim,R.B.
TITLE COMPLEMENT INHIBITOR
JOURNAL Patent: WO 9823638-A 4 04-JUN-1998;
Schwabele Wilhelm (GB); UNIV LEICESTER (GB)
FEATURES
Location/Qualifiers
1..1532
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 523 a 275 c 290 g 444 t
ORIGIN

Query Match 47.7%; Score 413.4; DB 6; Length 1532;
Best Local Similarity 99.8%; Pred No. 2.9e-98;
Matches 414; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tcgagtcactgtcccgatagatccagacatgagactgtcagcaagaattatttggc 60
Db 1 TCGAGTCAACTGCTCCAGATAGATCCAGACATGAGACTGTGACAGAAATATTGGC 60

Qy 61 ttatatattgactgttgttagcagaagattgtaaggctctctcccaagagaaaatt 120
Db 61 TTATATTATGGACTGTTTGTGTAGCAGAAGATTGTAAGGCTCTCTCCAGAGAAAATT 120

Qy 121 cagaattctctcaggttcgtctgaacaactatattcagaagcactcaggcaacct 180
Db 121 CAGAAATCTCTCAGGTTCGTGCTGTAACAACATATATTCAGAAAGGCATCGGCAACCT 180

Qy 181 acaaatgcgccttgataccgaacacttggtactattgttaaaagtatgcaagaatgag 240
Db 181 ACAAAATGCGCCTCGATACCAACACATGCTGACTATGTAAGATGCAAGATGGAG 240

Qy 241 aatgggtacctcttaaccatacaagatattcggaagagccatgtgggcatcccgag 300
Db 241 AATGGGTACCTTCTTAACCCATCAAGGATATGTCGAAAAGGCCATGTGGCATCCGGAG 300

Qy 301 acacaccttgggtcccttagctggcagttggtatctgaattgaaattggtgcaagg 360
Db 301 ACACACCTTTGGGTCCCTTAGCTGGCAGTTGGATCTGAATTTGAATTTGGTGCAGAGG 360

Qy 361 ttgtttacatgtgatgaagggtaccacactattagggtgaattgattaccgtga 415
Db 361 TTGTTTATACATGTGATGAAGGGTACCAACTATTAGTGAATTCATTACCGTTA 415

RESULT 9
LOCUS HSH38 1427 bp mRNA linear PRI 03-MAY-1999
DEFINITION Human mRNA for complement factor H 38-kDa N-terminal fragment.
ACCESSION X04697
VERSION X04697.1 GI:31991
KEYWORDS complement factor H.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1427)
Schulz,T.F., Schwabe,W., Stanley,K.K., Weiss,E. and Dierich,M.P.
AUTHORS Human complement factor H: Isolation of cDNA clones and partial
TITLE cDNA sequence of the 38-kDa tryptic fragment containing the binding
site for C3b
Eur. J. Immunol. 16 (11), 1351-1355 (1986)
JOURNAL 87054207
MEDLINE
COMMENT Clone H-19 codes for the first 108 AA of the 142-kDa fragment
in addition to the 289 AA derived from the 38-kDa fragment.
An additional A residue at pos. 930 not observed in a recently
published cDNA clone caused a shift in reading frame. It represents
not a sequencing artefact but is probably due to an error that

occurred during reverse transcription. Data kindly reviewed
(03-AUG-1987) by Schulz T.F.

FEATURES
Location/Qualifiers
1..1427
/organism="Homo sapiens"
/db_xref="taxon:9606"
misc_feature
1..867
/note="38-kDa C-terminal tryptic H fragment"
CDS
join(<1..929,931..1195)
/note="ORF"
/codon_start=1
/protein_id="CAB41739.1"
/db_xref="GI:4725976"
/translation="ILGYSLGNVIMVCKGEWALNPLRKQKRCRPGHPTGFTF
TLTGNNVEYGVKAVYTCNEGVLLEINRECDTDGNDIPICEVVKCLPVTAPEN
GKIVSSAMEPDREYHFGOAVFVCSGKIEGDEEMHCSDDGFWKEPKCEISCKS
PDVINGSPIKIIYKENERFOYKCMGYEYSESDGDAVCTESGWRPLSCSKSDNP
YIPNGDYLPHLRHRTGDEITTCRNGEYFATGNTAKTICWIPAPRCTLKPQDYP
DKHGGLYHENRRRYPFVAVGKYYSYCDHEFTFPGSYNDHIHCTODGWSPAVPC
LRCYFPYLENGYNQYGRKFVQGSIDVACHFGYALPKRAQTIVTTCMENGSPTPRCIR
VSFTL"
misc_feature
868..1192
/note="142-kDa N-terminal fragment"
misc_feature
930
/note="additional A residue, causing a shift in reading
frame"
polyA_signal
1407..1412
polyA_site
1427..1427
BASE COUNT 435 a 254 c 333 g 405 t
ORIGIN

Query Match 37.1%; Score 321.6; DB 9; Length 1427;
Best Local Similarity 78.0%; Pred No. 4.3e-74;
Matches 387; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 191 cctgtgataccgaacacttggtactattgtaaaagtatgcaagaatggaagatgggtacc 250
Db 3 CCTGGATATAGATCTCTTGGAAATGTAATAATGTTATGTCAGGAAGGAGAAATGGTTCG 62

Qy 251 ttctaaccatacaagatattcggaagagccatgtgggcatcccgaggagacacacctt 310
Db 63 TCTTAATCCATTAAGGAAATGTCAGAAAAGGCCCTGTGGACATCTCTGGAGATACTCTTT 122

Qy 311 tgggtcccttagctggcagttggatctgaattgaaattggtgcaagggtgtttatcac 370
Db 123 TGGTACTTTTACCCCTACAGGAGGAAATGTGTTGAATATGTTGTAAGAGCTGTGTATAC 182

Qy 371 atgtgatgaagggtaccacactattagggtgaattgattaccgtgaattgtatgcagatgg 430
Db 183 ATGTAATGAGGGGTATCAATTCCTAGGTGAGATTAAATACCGTGAATGTGACACAGATGG 242

Qy 431 gtggaccaatgatattccaataatggaattgtgaagtgcttgcagtgctgcagagaactgga 490
Db 243 ATGGACCAATGATATCTCTATATGTGAAGTTGTGAAGTTGTACCAGTGCACACACCA 302

Qy 491 gaatggaagaattgtgagtggtgcagccgaaccacagcaggaataattatttggacagtt 550
Db 303 GAATGGAATAATTTGTCAGTGTGCAATGAACCAATCGGGAATACCATTTTGGACAACG 362

Qy 551 ggtacgctttgaatccaactccggttcaagattgaaaggacagaagaataatcactgtctc 610
Db 363 AGTACGGTTGTATGTAACCTCAGGCTACAGATTCAAGAGATGAAGAAATGCATTGTTC 422

Qy 611 ataaaatgcctctgagcaaatgaaagccacagtggtggaatttcttgcctgccacc 670
Db 423 AGACATGGTTTTGGAGTAAAGAGAAAACCAAGTGTGTGGAAATTTTCATGCAAAATCCC 482

Qy 671 acgagttgaaatgga 686
Db 483 AGATGTTTAAATGGA 498

* 76781 79503: contig of 2723 bp in length
* 79504 79603: gap of unknown length
* 79604 82811: contig of 3208 bp in length
* 82812 82911: gap of unknown length
* 82912 85486: contig of 2575 bp in length
* 85487 85586: gap of unknown length
* 85587 88158: contig of 2572 bp in length
* 88159 88258: gap of unknown length
* 88259 90363: contig of 2105 bp in length
* 90364 90463: gap of unknown length
* 90464 93178: contig of 2715 bp in length
* 93179 93278: gap of unknown length
* 93279 95124: contig of 1846 bp in length
* 95125 95224: gap of unknown length
* 95225 98293: contig of 3069 bp in length
* 98294 98393: gap of unknown length
* 98394 101043: contig of 2650 bp in length
* 101044 101143: gap of unknown length
* 101144 104079: contig of 2936 bp in length
* 104080 104179: gap of unknown length
* 104180 105994: contig of 1815 bp in length
* 105995 106094: gap of unknown length
* 106095 108281: contig of 2187 bp in length
* 108282 108381: gap of unknown length
* 108382 110825: contig of 2444 bp in length
* 110826 110925: gap of unknown length
* 110926 113171: contig of 2246 bp in length
* 113172 113271: gap of unknown length
* 113272 115358: contig of 2087 bp in length
* 115359 115458: gap of unknown length
* 115459 117605: contig of 2147 bp in length
* 117606 118976: contig of 1271 bp in length
* 118977 119076: gap of unknown length
* 119077 121384: contig of 2308 bp in length
* 121385 121484: gap of unknown length
* 121485 123529: contig of 2045 bp in length
* 123530 123629: gap of unknown length
* 123630 125180: contig of 1551 bp in length
* 125181 125280: gap of unknown length
* 125281 126315: contig of 1035 bp in length
* 126316 126415: gap of unknown length
* 126416 127699: contig of 1284 bp in length
* 127700 127799: gap of unknown length
* 127800 129330: contig of 1531 bp in length
* 129331 129430: gap of unknown length
* 129431 130464: contig of 1034 bp in length
* 130465 130564: gap of unknown length
* 130565 132015: contig of 1451 bp in length
* 132016 132115: gap of unknown length
* 132116 134216: contig of 2101 bp in length
* 134217 134316: gap of unknown length
* 134317 135628: contig of 1312 bp in length
* 135629 135728: gap of unknown length
* 135729 137629: contig of 1901 bp in length
* 137630 137729: gap of unknown length
* 137730 139483: contig of 1754 bp in length
* 139484 139583: gap of unknown length
* 139584 141235: contig of 1652 bp in length
* 141236 141335: gap of unknown length
* 141336 142910: contig of 1575 bp in length
* 142911 143010: gap of unknown length
* 143011 144249: contig of 1239 bp in length
* 144250 144349: gap of unknown length
* 144350 145400: contig of 1051 bp in length
* 145401 145500: gap of unknown length
* 145501 147005: contig of 1505 bp in length
* 147006 147105: gap of unknown length
* 147106 148396: contig of 1291 bp in length
* 148397 148497: gap of unknown length
* 148497 150032: contig of 1536 bp in length
* 150033 150132: gap of unknown length
* 150133 151263: contig of 1131 bp in length.

FEATURES
source
1. .151263
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-76J6"
BASE COUNT 46018 a 28440 c 27866 g 43718 t 5221 others
ORIGIN
Query Match 15.4%; Score 133.4; DB 2; Length 151263;
Best Local Similarity 87.7%; Pred. No. 1.6e-24;
Matches 157; Conservative 0; Mismatches 21; Indels 1; Gaps 1;
Qy 689 tggatagagaaatcgagagatcacct-cggttcgtacaaagtgcattgagggtcac 747
|||||
Db 34914 TGGATATAGAAATTAAGAGGATCACCTCCCATTTGCTGTTCAAAATGCATTGATGTCACA 34855
Qy 748 tcaattatccactgtgtataaaatcgctatacaattattagtaaaccttatggatgac 807
|||||
Db 34854 TCAATTATCCACTGGTCTGTAAATACAAATGATTTATTAGTAAACCTTTATGGATCAA 34795
Qy 808 acttggttagaaatgcacatgtatattactaataacagtttgaaattacatttgaataa 866
|||||
Db 34794 CCTTTGTTAGAAATACACATGCTTATTACTAATTCATTTCAATTTTACATTTGAAATA 34736
RESULT 13
HSJ177P10/c
LOCUS
DEFINITION
Human DNA sequence from clone RPI-177P10 on chromosome 1q31.2-32.1
Contains 2 isoforms of the HFI (H factor 1 (complement)) gene,
ESTs, STSS and GSSs, complete sequence.
ACCESSION AL049744.8 GI:6782357
VERSION
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 150626)
AUTHORS
TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (26-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jan 27, 2000 this sequence version replaced gi:5531531.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Emi, EMBL; Swi, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep RPI-177P10 is
from the library RPI-1 constructed at the Roswell Park Cancer
Institute by the group of Pieter de Jong. For further details see
http://bacpac.med.buffalo.edu/
VECTOR: PCYPAC2
This sequence is the entire insert of clone RPI-177P10 The true
right end of clone RPI-15D12 is at 48650 in this sequence.
Location/Qualifiers

```

source
1. .150626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q31.2-32.1"
/clone="RPI-177P10"
/clone_lib="RPC1-1"
1841. .12140
repeat_region
repeat_region
misc_feature
5327. .5740
repeat_region
7345. .7430
repeat_region
7767. .8765
repeat_region
8750. .9063
repeat_region
9220. .10853
repeat_region
15437. .15809
repeat_region
15827. .16052
repeat_region
16050. .17506
repeat_region
17774. .18042
repeat_region
19407. .19722
repeat_region
19723. .20013
repeat_region
20014. .20732
repeat_region
24209. .24278
repeat_region
misc_feature
/note="H factor 1 (complement); match: cDNAs: Em:M12660
Em:X98697 Em:M29009 Em:M17517 Em:X07523 Em:X04697
Em:X68679 Em:M65293 Em:M65292 Em:X98337 Em:M29008
Em:X56210 Em:X56209 Em:M12383 Em:X64877 Em:M29007"
/product="dJ177P10.1.1 (H factor 1 (complement) isoform
1)"
/evidence=not_experimental
complement(24831. .120301)
/gene="HF1"
CDS
complement(join(25020. .25222,26334. .26516,28705. .28881,
30282. .30458,31541. .31714,34673. .34858,35327. .35509,
45393. .45572,45681. .45863,47036. .47212,56564. .56740,
58416. .58598,82094. .82270,82719. .82913,87096. .87269,
92540. .92710,94666. .94857,96268. .96344,98371. .98476,
99170. .99355,120158. .120301))
/gene="HF1"
gene
/note="H factor 1 (complement); match: cDNAs: Em:M12660
Em:X98697 Em:M29009 Em:M17517 Em:X07523 Em:X04697
Em:X68679 Em:M65293 Em:M65292 Em:X98337 Em:M29008
Em:X56210 Em:X56209 Em:M12383 Em:X64877 Em:M29007"
/product="dJ177P10.1.1 (H factor 1 (complement) isoform
1)"
/evidence=not_experimental
complement(24831. .120301)
/gene="HF1"
CDS
complement(join(25020. .25222,26334. .26516,28705. .28881,
30282. .30458,31541. .31714,34673. .34858,35327. .35509,
45393. .45572,45681. .45863,47036. .47212,56564. .56740,
58416. .58598,82094. .82270,82719. .82913,87096. .87269,
92540. .92710,94666. .94857,96268. .96344,98371. .98476,
99170. .99355,120158. .120215))
/gene="HF1"
/note="match: proteins: Sw:P06909 Sw:P08603 Tr:Q61408
Tr:Q02808 Tr:Q14570 Tr:Q61407 Tr:Q92496 Sw:Q02985
Tr:Q99254 Tr:Q91275"
/codon_start=1
/evidence=not_experimental
/product="dJ177P10.1.1 (H factor 1 (complement) isoform
1)"
/protein_id="CAB70597.1"
/db_xref="GI:6782358"
/translation="MRLAKIICMLWALICVAEDCNELPPRRNTEILTGSWDQTYPE
GTQAIYKCRPGYRLGNVIMVCRKGEWVALNPLRKCKORPCGHPGDTFGTFTLTGGN
VFYGVAVYTCNEGQQLLEINTRECDTGWINDIPICEVVKCLPVTAPNGKIVSS
AMPDREYHFGAOFYKCNMGYKTEGDEMHCSDDGFWSKPKCEIJSKSPDIVNG
SPIKOKLIYKNERFOYKCNMGYKTEGDAVCTESGWRPLPSCCEKSCDNPYIPNGD
YSLRLKHRTGDEITYOCNGFYPATRGNTAKCTSTGWIAPARCTLKPCDVPDIKHGG
LIHENMRPYFPAVAVGYIYSCDEHFPSPGSYHDIHICTQDGHSPAVPLCKICYFP
YLENGYNQHGKRFVQGSIDVACHPGYALPAKPAOTVTTCMENGWSPTRCIKRVKTSK
SSIDIENGFISESQTYTALKKAKYCKGLGYVADGETSGTITCGKDGMSAOPTCIKS
CDIPIVMNARTKNDFTWFLKNDITDYEDCHGYVESNTGTYKCFEGIDGAPICKE
RECLPKIDVHLVPDRKDKYKVEGLKFSCKPGFTIVGPNVOCYHFCGLSDLPICK
FOVOSCGPPPELNGNVEKTEKEYGHSVEVEYCNPRFLMKGPNKIOCVDGEWITLP
VCIYEESTCGDIPLEHGHGAQLSSPPYYIGDSVERNCSESTMHGHSITCIHGWTO
LPQCVAQIQLCPPPPQIPNSHNMTTTLNRYDGEKVSVLCCOENYLIQGEELITCKDGR
WOSIPLCVKEIKPCSQPQIEHGTINSSRSQESYAHGTKLSYTCGGFRISENETTC
YMGKWSPPQCEGLPCKSPPEISHGVVAHMSDSYQYGEVEYTKCFEGIDGAPICKE
LGEKWSHPSCIKTKDCLSLSPFENAIKPMGKKDKYKAGEQVITYTCATYKMDGASNV
CINSRWTCRPTCRDTSVNPPTVONAYIVSRQMSKYPSGERVRYOCRSPYEMFGDEEV
MCLANGWTEPPCKDSTGCKGPPPIPDNGDITSPFLSYVAPASSVEYQCONLYQLEGN
KRITCRNGWSEPPKCLHPCVISREIMENYIALRWTAQKLSYTSVEGSEVFCARGY
RLSSRSHTLRTTCDWGLKLEYPTCAK"
25148. .25518
/note="match: GSS: Em:AQ572291"
26837. .26998
/note="L1M4c repeat: matches 1281. .1443 of consensus"
27049. .28155
/note="L1M4c repeat: matches 1537. .2293 of consensus"
complement(28990. .29574)
/gene="HF1"
/note="match: GSS: Em:AQ322856"
29070. .29574
/note="match: GSS: Em:AQ309855"
29070. .29462
/note="match: GSS: Em:AQ015555"
32503. .34070
/note="THEIC-internal repeat: matches 1. .1580 of
consensus"
35512. .35633
/note="61 copies 2 mer aa 61% conserved"
36500. .36733
/note="117 copies 2 mer tt 54% conserved"
37623. .37757
/note="L1M3d repeat: matches 252. .385 of consensus"
37792. .37944
/note="L1M3d repeat: matches 508. .654 of consensus"
37955. .38048
/note="L1M3b repeat: matches -133. -.39 of consensus"
39598. .42996
/note="L1M1 repeat: matches 857. .3909 of consensus"
43390. .43463
/note="37 copies 2 mer ta 67% conserved"
43569. .43618
/note="25 copies 2 mer gt 94% conserved"
complement(43800. .44290)
/gene="HF1"
/note="match: GSS: Em:AQ595617"
44323. .44795
/note="match: GSS: Em:AQ785133"
complement(45113. .45337)
/gene="HF1"
/note="match: GSS: Em:AQ630668"
46395. .46801
/note="MSTB repeat: matches 4. .404 of consensus"
47551. .48221
/note="L1PBB repeat: matches -695. .29 of consensus"
48285. .48548
/note="L1PBB repeat: matches 82. .346 of consensus"
48556. .48620
/note="L1PBA repeat: matches -233. -.162 of consensus"
48674. .48853
/note="L1PAL5-16 repeat: matches 663. .848 of consensus"
48988. .50194
/note="L1M2 repeat: matches 764. .2360 of consensus"
50222. .50330

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 17:19:59 ; Search time 3911.83 Seconds
(without alignments)
2987.953 Million cell updates/sec

Title: US-09-316-163-2

Perfect score: 866

Sequence: 1 tcgagtcacactgtccccaga.....ttgaattacatttgaaaaa 866

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	496.6	57.3	652	9	BB609000
2	491.4	56.7	891	10	BF237071
3	448.4	51.8	842	9	AU122695
4	446.8	51.6	764	10	BI460026
5	446.8	51.6	906	9	AL540473
6	445.6	51.5	903	9	AL531827
7	442.8	51.1	890	9	AU116828
8	430.6	49.7	642	9	BB661130
9	401.2	46.3	722	10	BI760742
10	397.4	45.9	686	10	BG927932
11	378	43.6	988	10	BG288341
12	372.2	43.0	659	10	BG898846
13	370.8	42.8	650	9	A1207474
14	368.6	42.6	643	10	BG897905
15	367	42.4	662	10	BG900945
16	365.8	42.2	614	9	AL613807
17	355.8	41.1	640	10	BG901032

18	343	39.6	606	10	BG898574
19	341.4	39.4	578	9	AA500366
20	326	37.6	567	9	AW951533
21	296.2	34.2	395	9	BB845989
22	257.2	29.7	474	10	BF461587
23	252.8	29.2	760	10	BF027852
24	225.2	26.0	646	10	BG925340
c 25	209	24.1	525	9	AI683097
26	205	23.7	432	10	BG899280
27	201.4	23.3	411	10	BG898857
28	192.6	22.2	408	10	BG897147
29	186.6	21.5	743	9	AU117664
30	181	20.9	663	10	BG924822
31	179.8	20.8	398	10	BG898301
32	176.2	20.7	265	9	AA384423
33	176.6	20.4	377	10	BG897473
c 34	173.2	20.0	595	10	BF395144
c 35	173.2	20.0	626	9	AI009773
c 36	173.2	20.0	631	9	AI170314
37	172.4	19.9	649	10	BG925101
c 38	170	19.6	578	9	AI410851
c 39	168.2	19.4	484	9	AI009489
c 40	163	18.8	288	9	BB601574
41	148.8	17.2	423	10	BG926928
c 42	146	16.9	423	9	AA859019
c 43	142.8	16.5	364	9	AA945005
c 44	141	16.3	244	9	BB598679
45	139.6	16.1	300	9	AU098607

ALIGNMENTS

RESULT 1

BB609000

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB609000 RIKEN full-length enriched, 2 days pregnant adult female
ovary Mus musculus CDNA clone E33009L21 5', mRNA sequence.

BB609000.2 GI:16451137

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 652)

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hangaki,T., Hara,A.

, Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,K., Konno,H., Kouda

, M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,

Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki

, D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,

Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,

Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

On Dec 6, 2000 this sequence version replaced gi:11564176.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh

, M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

waghi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura

, S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and


```

QY 61 ttatatattgagcagctgttctgttagcagaagattgtaaaagctctctccaaagagaaatt 120
Db 94 TTATATATGGAGCT-TTGTGCGACAGACAGATTGTAAAGCTCCTCTCCAAGAGAAAAAT 152

QY 121 cagaataatcctcagctgtcgtggtctgaacaactatattcagaagcactcaggcaacct 180
Db 153 CAGAAATCTCTCAGGCTCGTGGTCAGACAACTATATCCAGNAGGCCACCCAGGCTACCT 212

QY 181 acaaatgcgcgcctggataccgaacacttggtactattgtataaagtatgcaagaatggag 240
Db 213 ACAAAATGCCGCCCTGGATACCGAACACTTGGCAGCTATTGTAAAGATGATCAAGAATGGAA 272

QY 241 aatgggtaccttctaaccatcaagatattgctggaaaaggccatgtggcctcccgag 300
Db 273 AATGGGTGGCGTCTAACCCATCAGGATATGCGGAAAAAGCCTTGTGGCATCCCGGAG 332

QY 301 acacaccccttgggtcctttagctggcagctggatctgaatttgaaattgattaccctggaatgtg 420
Db 333 ACACACCCCTTTGGGTCTTTAGGCTGGCAGTTGGATCTCAATTTGAGTTTGGTGCAAGG 392

QY 361 ttgtttacatcgtgatgaaggggtaccactatttagtgaaattgattaccctggaatgtg 420
Db 393 TTGTTTATACCTGTGATGATGGGTATCAACATATTAGTGAAATTTGATTACCGTGAATGTG 452

QY 421 atgcagatgggtggaccatattccaatatgtgaattgtgaagtgtcctgcagtgaa 480
Db 453 GTGCAGATGGGTGGATCAATGATATTCCACTATGTGAAGTTGTGAAGTCTCTACCTGTGA 512

QY 481 cagaactgagaatggaagaattgtgagtggtgagccgaaccagaccaggaatattatt 540
Db 513 CAGAATCTCAGAAATGGAAGATCGTGAGTGGTGCAG-AGAAACAGACACGAGGAATACTATT 571

QY 541 ttggacaggtggtacgcttcttgatgcaactccggtcctcaagattggaag 589
Db 572 CAGGACAGCTGGTGGGTTTGAATGCAATTTCAGGCTTCAAGATTGAAG 620

RESULT 3
LOCUS AUI22695 MAMMAL Homo sapiens cdna clone MAMMAL1002920 5', mRNA EST 23-OCT-2000
DEFINITION AUI22695 MAMMAL Homo sapiens cdna clone MAMMAL1002920 5', mRNA
sequence.
ACCESSION AUI22695
VERSION AUI22695.1 GI:10947411
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yanamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cdna project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cdna project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cdna library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. .842
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MAMMAL1002920"
/clone_lib="MAMMAL"

FEATURES
source

```

```

/issue_type="mammary gland"
/note="Vector: pME18SFL3"
BASE COUNT 278 a 124 c 205 g 232 t 3 others
ORIGIN

Query Match 51.8%; Score 448.4; DB 9; Length 842;
Best Local Similarity 78.6%; Pred. No. 2.3e-111;
Matches 536; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 5 gtcaactgctccagatagatccaacatgacacgtgacacgtgacgaagaattatttgcttat 64
Db 46 GTPAATGTCTCTTAAAGATCCAAAATAGACACTTCTAGCAAAAGATTATTTCCTTAT 105

QY 65 attatggaactgttctgtagcagaagattgtaaaaggtccctccccaagagaaattcaga 124
Db 106 GTTATGGGCTATTGTGTAGCAGACAGATTGCAATGAACCTTCCTCCAAGAGAAATACAGA 165

QY 125 aattctcaggtctggtctgtaacaactatattcagaaggcactcaggaacacctacaa 184
Db 166 AATTCCTGACAGGTTCTCGTGTCTGACCAACATATCCAGAAGGCCACCCAGGCTATCTATAA 225

QY 185 atgcccgcctgataccgaacacttggtactattgtaaaagtgtggaagtatgcaagaatgagaatg 244
Db 226 ATGCCCCCTGGATATGATCTCTTGGAAATATAATGTTGTATGCGAAGGAGGAATG 285

QY 245 ggtaccttctaaacctcaaggatattgctggaagggccatgtggcctcccgagacac 304
Db 286 GGTGTCTCTTAATCCATTAAAGGAATGTCAGAAAGGCCCTGTGGACATCTCTCGAGATAC 345

QY 305 accttgggtccctttagctggcagctggtgactgtgaatttgatggcgaaggtgtg 364
Db 346 TCGTTTGTGTTACTTTTACCCCTTACAGGAGGAATGTGTTTGAATATGTTGTAAGCTGT 405

QY 365 ttatcatgtgatgaaggtaccacactattagtgaaattgattaccggaagtgtgacg 424
Db 406 GTATACATGTAATGAGGGGTATCAATGCTAGGTGAGATTAATACCGTGAATGTGACAC 465

QY 425 agatgggtggaaccaatgatattccaatatgtgaagtgtgaaagtgtgcaagctgtgcagtgacaga 484
Db 466 AGATGGATGGACCAATGATATTCTATATGTGAAGTTGTGAAGCTGTTTACCAGTGACAGC 525

QY 485 actggagaatggaagaattgtgagtggtgagccgcagccgaaccagaccaggaatatattttgg 544
Db 526 ACCAGAGAATGGAAAAATTTGTCAGTAGTGCATTAATGCAACCAAGATCGGGAATACCATTTTG 585

QY 545 acaggtggtacgcttgaatgcaactccgcttccaagattggaagtgacagacagaagaatgca 604
Db 586 ACAAGCAGTACGGTTTGTATGTAACTCAGGCTACAAAGATTGAAGGAGATGAAGAAATGCA 645

QY 605 ctgctcataaaatggcctctggagcaaatgaaaagccacagtggtgtggaattttcttgct 664
Db 646 TTGTTTCAGACGATGGTTTTCGAGTAAACAGAGAAACCAACCAAGTGTGTGGAATTTTCATGCA 705

QY 665 gcaacacaggtggaataatgga 686
Db 706 ATCCCCAGATGTTATATAAATGGA 727

RESULT 4
LOCUS BI460026
DEFINITION BI460026 603201565F1 NIH_MGC_97 Homo sapiens cdna clone IMAGE:5267249 5',
mRNA sequence.
ACCESSION BI460026
VERSION BI460026.1 GI:15250682
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 764)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

```

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs-r@mail.nih.gov Tissue procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Pietro Carninci (RIKEN) cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLMI1674 row: b column: 18 High quality sequence start: 10 High quality sequence stop: 764.

```

FEATURES
  source
    Location/Qualifiers
      1..764
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_xref="IMAGE:5267249"
        /clone_lib="NIH_MGC_97"
        /lab_host="DH10B"
        /note="Organ: testis; Vector: pBluescript (modified
        pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gcgcag
        ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTTTVN-3',
        size-selected for average insert size 2.2 kb and
        normalized to 10^5. This is a primary library enriched
        for full-length clones and constructed using the
        Cap-trapper method (Carninci, in preparation). Library
        constructed by M. Brownstein (NIMH/NHGRI, National
        Institutes of Health). Note: this is a NIH_MGC Library."
      254 a 117 c 185 g 208 t

```

Query Match	51.6%	Score	446.8;	DB	10;	Length	764;
Best Local Similarity	78.4%;	Pred.	No. 5.9e-111;				
Matches	535;	Conservative	0;	Mismatches	147;	Indels	0;
QY	5	gtcaactgctcccgatagatcacaagaacatgcagactgcagaagaataattttggcttat	64				
Db	45	GTAAATGTCCTCTTAAAGATGCCAAAATGCACACTTCTAGCAAGAAGATTATTTGCCCTAT	104				
QY	65	attatgaactgttctgttagcagaagtgttaaaggctccctcccacaagaaaaattcaga	124				
Db	105	GTTATGGGCTATTGTGTACAGAAGATTGCAATGAACCTTCTCCCAAGAANAATCACA	164				
QY	125	aattctctcaggttcgttggtctgaacaactatatcagaaggcactcagcgacacctcaa	184				
Db	165	AATTCTGACAGGTTCTCTGGTCTGACCAAACATATCCAGAAGGCACCAGGCTATCTATAA	224				
QY	185	atgccgccttggtatccgaacaatttgtaactatgttaaaagtatgcagaagaatggaaatg	244				
Db	225	ATGCCGCCCTGGATATAGATCTCTTGGAATGTAAATAATGCTATCGAGGAAGGAGAATG	284				
QY	245	ggtacctcttaaccocatcagaagatatgtcggaaaaggccatgtgggcatcccgagagacac	304				
Db	285	GGTGTCTTTAATCCATTAGGAAGATGTGAAAAAGCCCTGTGGACATCTCTGGAGATAC	344				
QY	305	accctttgggtccttlagggctgcagttggaatcgtgaatttgaatttgggtcgaaggttgt	364				
Db	345	TCCCTTTGGTACTTTTACCCTTACAGGAGGAATGTGTTTGAATATGGTGTAAAGCTGT	404				
QY	365	tatacatgtgatgaagggttaccaactattaggatgaattgattaccgtgaatgtgatgc	424				
Db	405	GTATACATGTAAATGAGGGGTATCAATTGCTAGCTGAGATTAATACCGTGAATGTGNACAC	464				
QY	425	agaatgggtggaccaatgatattccaatatgtgaagttgtgaagtcttgcgcagtgacaga	484				
Db	455	AGATGTGATGGACCAATCATATTCCTATATGTGAAGTTGTGAAGTGTTCACAGTGACAGC	524				
QY	485	actggagaatggaagaattgtgaatgggtgcagccgaaccagaccagaagaatatattttgg	544				

[illegible]

```

FEATURES
source
Location/Qualifiers
1. .906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS00E001XB21"
/clone_lib="LTI_FL002_PL1"
/lab_host="DHI0B"
/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fullenatb.invitrogen.com"

```

```

BASE COUNT
ORIGIN
299 a 138 c 214 g 255 t

```

[illegible]

```

210 ATGCCGCCCTGGATATAGATCTCTTGGAAATGTAATATGGTATGAGGAAGGAGAATG 269
QY 245 ggtacaccttaaccatcaagcatatgctcgaaagccatgtggcattcccgagagacac 304
Db 270 GGTGCTCTTAATCCATTAAAGNAATGTCAGAAAAGGCCCTGTGGACATCCTGGAGATAC 329
QY 305 accctttgggtcccttaggctggcagtgatggaattgaatttggtgcaagattgt 364
Db 330 TCCTTTTGGTACTTTTACCCCTTACAGGAGGAATGTGTTTGAATATGGTCTAAAGCTGT 389
QY 365 ttatacatgtgatgaaggtaccacattatagtgaaattgattaccgtgaattgtatgc 424
Db 390 GTRATACATGTAATGACGGGTATCAATGCTAGGTGAGATTAATTACCGTGAATGGTGCAC 449
QY 425 agatgggtggaccaatgatattccaatatgtgaagttggaagtgtgtccagtgacaga 484
Db 450 AGATGATGACCAATGATATTCCTATATGTGAAGTTGTGAAGTGTACCAGTGACAGC 509
QY 485 actggagaatgaagaattgtgagtggtgcagccgaaccagaccaggaatatattttgg 544
Db 510 ACCAGAGAATGGAATAATTTGTAGTGTGCAATGGAACCCAGATCGGGAATACCATTTTGG 569
QY 545 acaggtggtacgtttggaatgcaactccgcttcaagattgaaggtggaagacagaagaatgca 604
Db 570 ACNAGCAGTACGGTTGTATGTAACTCAGGTACAGATTTGAAGGAGATCAAGAAATGCA 629
QY 605 ctgctcataaaatggcctctggagcaatgaaagccacagtggtgtggaattcttgcct 664
Db 630 TTGTTTCAGACGATGGTGTGTTTGGAGTAAAGAGAAACCAAAAGTGTGTGGAATTTTCATGCAA 689
QY 665 gccaccacagagttgaaatgga 686
Db 690 ATCCCCCAGATGTTATAAATGGA 711

RESULT 6
AL531827 6
LOCUS AL531827 LTI_NFL001_NBC4 903 bp mRNA linear EST 13-FEB-2001
DEFINITION AL531827 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DM003YM12 5
prime, mRNA sequence.
ACCESSION AL531827
VERSION AL531827.1 GI:12795320
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..903
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DM003YM12"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
```

```

8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com*
BASE COUNT 299 a 139 c 204 g 253 t 8 others
ORIGIN
Query Match 51.5%; Score 445.6; DB 9; Length 903;
Best Local Similarity 78.2%; Pred. No. 1.3e-110;
Matches 533; Conservative 2; Mismatches 147; Indels 0; Gaps 0;
QY 5 gtcaactgtccccagatagatccaagaatgagactgtcagcaagaattatttggcttat 64
Db 19 GTAATGTCTCTTAAAGATCCAAAATGAGACTTCTAGCAAAAGATTATTTCCTTAT 78
QY 65 attatggaactgtttgttagcagaagattgttaaaggctctccctcccaagagaaaattcga 124
Db 79 GTTATGGCTATTGTGTACGACAAGATTCGAATGAACCTTCTCCCAAGAGAAATACAGA 138
QY 125 aattctctcaggttcgtggtctgaaacaactatatcagaaggcactcaggcaacctacaa 184
Db 139 AATTCTGACAGGTTCTGTGCTGACCAACATATCCAGAAGGCCACCCAGGCTATCTATAA 198
QY 185 atgcgccttgataccaagaacacttggtactattgtaaaagtatgcaagaatggagaatg 244
Db 199 ATGCCGCCCTGGATATAGATCTCTTGGAAATGTAATATGGTATGCDGGRAGGAGAATG 258
QY 245 ggtacaccttaaccatcaaggtatgtcgaaagccatgtggcattcccgagagacac 304
Db 259 GGTGCTCTTAATCCATTAAAGNAATGTCAGAAAAGGCCCTGTGGACATCCTGGAGATAC 318
QY 305 accctttgggtcccttaggctggcagtgatggaatttggtgcaagattgtgcaaggtgtg 364
Db 319 TCCTTTTGGTACTTTTACCCCTTACAGGAGGAATGTGTTTGAATATGGTGTAAAGCTGT 378
QY 365 ttatacatgtgatgaaggtaccacattatagtgaaattgattaccgtgaattgtatgc 424
Db 379 GTRATACATGTAATGAGGGTATCAATGCTAGGTGAGATTAATTACCGTGAATGTGACAC 438
QY 425 agatgggtggaccaatgatattccaatatgtgaagttgtggaagtgtgtccagtgacaga 484
Db 439 AGATGATGACCAATGATATTCCTATATGTGAAGTTGTGAAGTGTACCAGTGACAGC 498
QY 485 actggagaatgaagaattgtgagtggtgcagccgaaccagaccaggaatatattttgg 544
Db 499 ACCAGAGAATGGAATAATTTGTAGTGTAACTCAGGTACAGATTTGAAGGAGATGAAGAAATGCA 558
QY 545 acaggtggtacgtttggaatgcaactccgcttcaagattgaaggtggaagacagaagaatgca 604
Db 559 ACNAGCAGTACGGTTGTATGTAACTCAGGTACAGATTTGAAGGAGATGAAGAAATGCA 618
QY 605 ctgctcataaaatggcctctggagcaatgaaagccacagtggtgtggaattttctgcct 664
Db 619 TTGTTTCAGACGATGGTGTGTTTGGAGTAAAGAGAAACCAAAAGTGTGTGGAATTTTCATGCAA 678
QY 665 gccaccacagagttgaaatgga 686
Db 679 ATCCCCCAGATGTTATAAATGGA 700
```

RESULT 7

```

ALU116828 7
LOCUS ALU116828 HEMBA1 Homo sapiens cDNA clone HEMBA1000088 5', mRNA
DEFINITION ALU116828 sequence.
ACCESSION ALU116828
VERSION ALU116828.1 GI:10931750
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 890)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
```

Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
Location/Qualifiers
1. .890
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1000088"
/clone_lib="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/note="Vector: pME18SFL3"
BASE COUNT 298 a 136 c 213 g 239 t 4 others
ORIGIN

Query Match 51.1%; Score 442.8; DB 9; Length 890;
Best Local Similarity 78.3%; Ref. No. 7.7e-110;
Matches 531; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 5 gtaactgtctccagatagatccaagacatgagactgtcagcaagaattatttggcttat 64
Db 49 GTAATGTCCTCTTANAAGATCCAAAATGACACTCTAGCAAAATATTATTGCCCTTAT 108
QY 65 attatgactgtttgtgtagcagaagattgtaaagtctctctcccaagagaaattcaga 124
Db 109 GTTATGGGTATTGTGTAGCAGAAGATTGCAATGAACCTTCTCCCAAGAGAATAACAGA 168
QY 125 aattctctcaggctgtgtctgaacaactattatcagaagcaactcaggcaacctacaa 184
Db 169 AATTCTGACAGGTTCTGTGTGACCAACAACATATCCAGAAAGGCCACCGCTATCTATAA 228
QY 185 atgccgccttggtataccgaacacttggtactattgttaaagtatgcaagaatggagaattg 244
Db 229 ATGCCGCCCTGGATATAGATCTCTTGGAAATATGATAATGGTATGTCAGGAAGGGAATC 288
QY 245 ggtaccttctaaccataaagatatgtcggaaaaggccatgtgggcatcccgagacac 304
Db 289 GGTTGCTCTTAATCCATTAAAGGAATGTGAGAAAGGCCCTGTGGACATCCTGGAGATAC 348
QY 305 accctttgggtccctttaggctggcagttggtatctgaattggaattgggtgcaaaagtgtt 364
Db 349 TCCTTTTGTACTTTTACCCTTACAGGAGAAATGTTTGAATATGTTGTTAAACACTGT 408
QY 365 ttatacatgtgatgaagggtacaaactatagtgtaaatgattaccgtgaattgtatgc 424
Db 409 GTATACATGTAATGAGGGGTATCAATTTGCTAGGTGAGATTAATTACCGTGAATGTGACAC 468
QY 425 agatgggtgaccaatgatattccaatatgtgaagtgtgaagtgtccagtgacaga 484
Db 469 ACATGATGACCAATGATATTCCTATATGTGTAGTGTGAAGTGTTTTACCAGTGACAGC 528
QY 485 attggagaatggaagaattgtgagtggtgagcccgacccagaccaggaataattatttgg 544
Db 529 ACCAGAGAAATGAAAAATTTTCAGTAGTCAATGGAACACAGATCGGAATATCAATTTGCG 588
QY 545 acaggtgttaactgtgaatgaactccggcttcaagattggaagacagagaagaatgca 604
Db 589 ACAAGCAGTACGTTGTGTGTAATCTACCTACGCTACAAAGATTGAAGGATGAAGAAATGCA 648
QY 605 ctgtctcaataaattggcctctggagcaatgaaagccacagtggtgtggaatttcttgcct 664

Db 649 TTGTTCCAGACGATGGTTTTTGGAGTAAGAGAACCAACGATGCTGCGAAATTCATCA 708
QY 665 gccaccacagagttgaaaa 682
Db 709 ATCCCCAGATGTTATATA 726

RESULT 8
BB661130
LOCUS
DEFINITION
musculus cDNA clone D630023013 5', mRNA sequence.
ACCESSION
BB661130
VERSION
BB661130.1
KEYWORDS
GI:16494909
SOURCE
EST.
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 642)
REFERENCE
AUTHORS
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda,
M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki,
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T.,
Muranatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayata, N., Sugahara, Y., Shibata, K., Itoh,
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura,
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara,
Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yananaka, I., Aizawa,
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
Hayashizaki, Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.
Location/Qualifiers
1. .642
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D630023013"
/clone_lib="RIKEN full-length enriched, 0 day neonate
kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"


```

QY 489 gagaatggaagaattgtgagtggtgcagccgaacacagaccagggaataattattttggacag 548
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 521 GAGAAATGGAATAATGTGCTAGTAGTGCATGGAACAGATCGGGAATACCATTTTGGACAA 580
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 549 gtggtacgctttgaatcaactcgcgttcagattgaagacagagaagaataatgcactgc 608
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 581 CGATGACGGTTGTATGTAAGTACAGCTACAGGATACAGGATGACGGAGATGAAGAAATGCATG 640
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 609 tcataaatgacct-citggagcaatgaaagccacagctgtgtggaattt 657
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 641 TCACACGATGGTTGATGGAGTAGAAGAGAAACCAAGTGTGTGGAACATT 690
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
BG927932 686 bp mRNA linear EST 06-NOV-2001
LOCUS HNC66-1-H3.R HNC (Human Normal Cartilage) Homo sapiens cDNA, mRNA
DEFINITION
sequence.
ACCESSION BG927932
VERSION BG927932.1 GI:14322455
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 686)
Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
Sathe,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
Lark,M.W.
IDENTIFICATION AND INITIAL CHARACTERIZATION OF 5000 EXPRESSED
SEQUENCED TAGS (ESTs) EACH FROM ADULT HUMAN NORMAL AND
OSTEOARTHRIC CARTILAGE CDNA LIBRARIES
Osteoarthritis Cartilage 9 (7), 641-653 (2001)
21482651
Contact: Sanjay Kumar
UW2109
GlaxoSmithKline
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
Tel: 610-270-7245
Fax: 610-270-5598
Email: sanjay_kumar-legsk.com
Seq primer: T7.
Location/Qualifiers
1. .686
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="cartilage"
/lab_host="E.coli DH10 B"
/notes="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
Directional"
BASE COUNT 214 a 126 c 167 g 179 t
ORIGIN
Query Match 45.9%; Score 397.4; DB 10; Length 686;
Best Local Similarity 79.5%; Pred. No. 1.6e-97;
Matches 470; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 2 cgaagtcaactgcctccagatagatccaagacatgagactgcagcaagaattatttgct 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94 CGGGTAAATGCTCTCTTAAAGATCCAAAATGAGACTTCTAGCAAGATTATTGGCT 153
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 62 tatattatgactgtttgttagacaagattgttaaggtcctccacagaaaaattc 121
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154 TATGTTATGGGCTATTGTGTATGACAGAAGATTGCAATGAATCTCTCCCAAGAAAGATAC 213
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 122 agaaattctctcaggtcgtcggtcgaacactatattcagaagggcactcagcaacct 181
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 214 AGAAATCTCACAGGTTCTGGTCTGACCAACATATCCAGAAGGCCACCGGCTATCTA 273
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 182 caaatgccccctggatccgaacacttggtactattgtataaagtatgaagaatgaga 241
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 274 TAAATGCCGCCCTGGATATAGATCTCTTTGAAATATAATAATGTTATGCAGGAAGGAGA 333
QY 242 atgggtacctctaccatcaaggatgtcggaaaaagccatgtggcattcccgaga 301
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 334 ATGGTTGCTCTTAATCAAGAAATGTCAGAAAAAGGCCCTGTGGACATCTCCGGAGA 393
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 302 cacaccctttgggtccttttagctggcagttggactcctgaatttgaatttgggtgcaaggt 361
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 TACTCCTTTTGGTACTTTTACCCTTACAGGAGGAATCTGTTTGAATATGGTGTAAAGC 453
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 362 tgtttatcatgtgatgaagggtaccactattaggtgaaattgattccctggaatgga 421
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 TGTGTATACATGTAATGAGGGGTATCAATGCTAGGTAGGATTAATATCCGTGAATGTGA 513
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 422 tgcagatgggtggacccaatgatattccaatatgtgaagtgtgaagtgtcgtccagtgac 481
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 514 CACAGATGGATGGACCAATGATATCTTATATGTGAAGTGTGGAAGTGTTCACCAAGTAC 573
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 482 agaactggagaatggaagaattgtgagtggtgcagccgaacacagaccagggaataattatt 541
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 AGCACCAGAGAATGGAATAATTTGTCAGTAGTGCATGATGCAATGCAACAGATCGGAATACCA 633
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 542 tggacaggtgtacgtcttgaatgcactcgcgttcgaatttgaatttgaaggaga 592
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 634 TGGACAAGCAGTACGGTTTGTATGTAAGTGTGCAAGGCTTACCAAGATTGAAGGAGA 684
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
BG288341 988 bp mRNA linear EST 21-FEB-2001
LOCUS 602383678F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4512906 5',
DEFINITION mRNA sequence.
ACCESSION BG288341
VERSION BG288341.1 GI:13043080
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 988)
NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLMI0398 row: k column: 19
High quality sequence stop: 661.
Location/Qualifiers
1. .988
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4512906"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/notes="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 336 a 134 c 259 g 258 t 1 others
ORIGIN
Query Match 43.6%; Score 378; DB 10; Length 988;
Best Local Similarity 76.2%; Pred. No. 3.6e-92;

```

Matches 491; Conservative 0; Mismatches 150; Indels 3; Gaps 2;

```

QY 5 gtaactgtctccagatagatccaaagacatgagactgtcagcaagaattatttggcttat 64
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 28 GTAATGTCTCTTAAAGATCCAAAATGAGACTTCTAGCAAAAGATTATTTCCTTAT 87
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 65 attagactgttgtgtagcagaagattgaaggctctccccaagagaaaaattcaga 124
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 88 GTTATGGGCTA-TTGTGTACGAGAAGATTGCAATGAAGTCTTCCCAAGAGAATAACAGA 146
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 125 aattctctcaggctcgtggtctgaaacaactatttcagaaggcactcaggcaacctacaa 184
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 147 AATTCTGACAGGTTCTGTGGTCTGACCAACATATCCAGAAGGCCAGCCAGGCTATCTATAA 206
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 185 atgccccttggtatccgaacaacttggtactattgttaaaagtatgcaagaatggagaatg 244
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 207 ATCCGCCCTGGATATAGATCTCTTGGAAATGTAATAATGCTATGCAAGGAGGAGAATG 266
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 245 ggtacctcttaacccatcaggatattgtcgaaagccatgtggcctcccgagacac 304
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 267 GGTGTCTTTAATCCATTAAGGAATGTCAGAAAGGCCCTGTGGACATCCTGGAGATAC 326
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 305 acccttgggtcctttagctggtcagtgatctgaaatttgaatttgggtcaagggtgtg 364
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 327 TCCTTTTGTGTACTTTACCTTACAGGAGGAATGCTTTTGAATATGGTGTAAAGCTGT 386
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 365 ttatacatgtgatgaagggtaccacactatttagtgaaattgattaccgtgaattgtatgc 424
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 387 GTATAGATGTAATGAGGGGTATCAATTTGCTAGGTGAGATTAAATACCGTGAATGTGACAC 446
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 425 agatgggtgagaccaatgatattccaatatattgaatttgaatttgggtcagtgacaga 484
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 447 AGATGATGAGCAACATGATATCTTATGTGAAGTTGTGAAGTGTGTACCACTGACAGC 506
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 485 actggagaatggaagaattgtgagtggtgcagccgaacacagaccaggaattattttgg 544
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 507 ACCAGAGAATGGAAAAATGTCAGTAGTGAATGGAACACAGATCGGGAATACCATTTGGA 566
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 545 acagggtggtacgtttgaaatgcaactcggctcagattgaaggacagaaagaatgca 604
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 567 CAAGCAGTAGCGG--TTGTATGTAACTCAGGCTAAAGATTGAAGGAGATGAAGAAATGCA 624
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 605 ctgctcataaattgctctgagcaaatgaaagccacagtggtg 648
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 625 TTGTTTACAGCATGGTTTGGAGTAAGAGAAACCAAGTGTGTG 668
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 12

LOCUS BG988846 659 bp mRNA linear EST 06-NOV-2001
 DEFINITION HOA4-1-D-8 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, mRNA sequence.

ACCESSION BG988846
 VERSION BG988846.1 GI:14309095
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 659)
 Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
 Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
 Lark,M.W.

TITLE Identification and initial characterization of 5000 expressed
 sequenced tags (ESTs) each from adult human normal and
 osteoarthritic cartilage cDNA libraries

JOURNAL Osteoarthritis Cartilage 9 (7), 641-653 (2001)

MEDLINE 21482651

COMMENT Contact: Sanjay Kumar

UW2109

GlaxoSmithKline
 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
 Tel: 610-270-7245

Fax: 610-270-5598
 Email: sanjay_kumar-1@sk.com
 Seq primer: T7.
 FEATURES
 source 1..659
 Location/Qualifiers
 1..659
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HOA (Human Osteoarthritic Cartilage)"
 /tissue_type="cartilage"
 /lab_host="E.coli DH10 B"
 /note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI;
 Directional"
 BASE COUNT 204 a 116 c 162 g 177 t
 ORIGIN

Query Match 43.0%; Score 372.2; DB 10; Length 659;
 Best Local Similarity 79.0%; Pred. No. 1.2e-90;
 Matches 443; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```

QY 5 gtaactgtctccagatagatccaaagacatgagactgtcagcaagaattatttggcttat 64
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98 GTAATGTCTCTTAAAGATCCAAAATGAGACTTCTAGCAAAAGATTATTTCCTTAT 157
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 65 attagactgttgtgtagcagaagattgtaaaaggctctccccaagagaaaaattcaga 124
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 158 GTTATGGGCTATTGTGTAGCAGAAGATTGCAATGAAGTCTTCCCAAGAGAATAACAGA 217
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 125 aattctctcaggctcgtggtctgaaacaactatttcagaaggcactcaggcaacctacaa 184
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 218 AATTCGACAGGTTCTGTGGTCTGACCAACATATCCAGAAGGCCAGCCAGGCTATCTATAA 277
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 185 atgccccttggtatccgaacaacttggtactattgttaaaagtatgcaagaatggagaatg 244
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 278 ATCCGCCCTGGATATAGATCTCTTGGAAATGTAATAATGCTATGCAAGGAGGAGAATG 337
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 245 ggtacctcttaacccatcaggatattgtcgaaagccatgtgggcatcccgagacac 304
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 338 GGTGTCTTTAATCCATTAAGGAATGTCAGAAAGGCCCTGTGGACATCCTGGAGATAC 397
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 305 acccttgggtcctttagctggtcagtgatctgaaatttgaatttgggtcgaagggtgtg 364
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 398 TCCTTTTGTGTACTTTTACCCTTACAGGAGGAAATGTGTTGAATATGTTGTAAGGCTGT 457
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 365 ttatacatgtgatgaagggtaccacactatttagtgaattgattaccgtgaattgtgatgc 424
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 458 GTATACATGTAATGAGGGGTATCAATTGCTAGGTGAGATTAAATTCCTGCTGAATGTGACAC 517
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 425 agatgggtggaaccaatgatattcccaatattgtgaagtgtgaaagtgttgcagtgacaga 484
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 518 AGATGGATGGACCAATCATATCTCTATATGTGAAGTGTGTGAAGTGTGTACCAAGTGACAGC 577
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 485 actggagaatggaagaattgtgagtggtgcagccgaacacagaccaggaataattatttgg 544
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 578 ACCAGAGAATGGAAAAATTTGTCAGTAGTGCATGAATGGAAGGAGATCGGGAATACCATTTGG 637
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 545 acagggtgtacgcttttgaatg 565
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 638 ACAAGCAGTACCCTTTGTATG 658
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 13

LOCUS AI207474 650 bp mRNA linear EST 11-NOV-1999
 DEFINITION HA2827 Human fetal liver cDNA library Homo sapiens cDNA, mRNA sequence.
 ACCESSION AI207474
 VERSION AI207474.1 GI:6361482
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 650)
AUTHORS Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.
          and He,F.
TITLE Expression profile analysis of a human fetal liver cDNA library
JOURNAL Unpublished (1998)
COMMENT Contact: Chenggang Zhang
          Beijing Institute of Radiation Medicine
          27 Taiping Road, Beijing 100850, P.R.China
          Email: zhang_chenggang@hotmail.com.

FEATURES
source Location/Qualifiers
          1..650
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_lib="Human fetal liver cDNA library"
          /tissue_type="liver"
          /dev_stage="fetal"
          /lab_host="MC1061/p3"
          /note="Vector: pCDNA1"

BASE COUNT 208 a 102 c 158 g 182 t
ORIGIN

Query Match 42.8%; Score 370.8; DB 9; Length 650;
Best Local Similarity 76.7%; Pred. No. 2.8e-90;
Matches 480; Conservative 0; Mismatches 142; Indels 4; Gaps 2;

QY 5 gtaactgtctccagatagatcccaagacatgagactgtcagcaagaattatttggcttat 64
DB 18 GTAATGTCCTTTAAAAGATCCAAAATAAGAGACTTCTAGCAAGATTTATTCCTTAT 77

QY 65 attatgactgtttgttagcagaagattgttaaaggctctctcccaagagaaattcaga 124
DB 78 GTTATGGCTATTTGTGTAGCAGAAGATTGCAATGAACTTCTCTCCCAAGAGAAATACAGA 137

QY 125 aattctctcagggttcgtgctgaacaactatttcagaagcactcaggcaacctcaaa 184
DB 138 AATTCTCACAGGTTCTGTGGTCTGACCAACATATCCAGAAGGCCACCCAGGCTATCTATA 197

QY 185 atgcgcgcctggataccgaacacttggtactattgttaaaagtatgcagaatggagaatg 244
DB 198 ATGCGCGCCTGGATATAGATCTCTTGGAAATGTAATAATGATGTCAGGAAGGGAGAATG 257

QY 245 ggtacctctcaaccatcaaggatattcggaagaagccatgtggcattcccgagagacac 304
DB 258 GGTGTCTCTTAATCCATTAGGAATAATGTCAGAAAGGCCCTGTGGACATCTCGAGATAC 317

QY 305 acccttgggtccctttaggtggcagttggatctggaattggaattgtgtgcaaaaggtgtg 364
DB 318 TCCTTTTGGTACTTTTACCCTTACAGCAGGAAATGCTGTTGAATATGTTGTAAGGCTGT 377

QY 365 ttatactgtgataagggtaccactatttagtgaaattgattaccgtgaatgtgatgc 424
DB 378 GTATACATGTAATGAGGGGTATCAATTTGCTAGTGTGAGATTAAATTTACCGTGAATGTGAC 437

QY 425 agatgggtggaccaaa----tgattccaatatgtgaaagttgaaagttgaaagttgccagtgc 481
DB 438 AGATGGATGTGACCAATGATATATCTCTATATGTGAAGCTGCGAAGTGTTTACCGTGAC 497

QY 482 agaactggagaaatggagaattgtgagtgagcgcggaaccagaccaggaattattatt 541
DB 498 AGTACCAGAAATGGAAAAATTTGTCAGTCTGTCGAATGGAACCCAGATCGGGAATACCAATT 557

QY 542 tggacaggtggtacg-ctttgaatgaactccggcttcaagattgaaggacagaaagaaa 600
DB 558 TGGACAAAGCAGTACGTGTTTGTATGTAACTCAGGCTACAAAGATTGAAGAGAGATGAAGAAA 617

QY 601 tgcactgctcataaaatggcctctgg 626
DB 618 TGCATTGGTCAGACCATGGTTTTTTG 643

RESULT 14
BG897905

```

```

LOCUS BG897905
DEFINITION HOA26-1-84 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
          mRNA sequence.
ACCESSION BG897905
VERSION BG897905.1 GI:14308154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 643)
AUTHORS Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
          Sathe,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
          Lark,M.W.
TITLE Identification and initial characterization of 5000 expressed
          sequenced tags (ESTs) each from adult human normal and
          osteoarthritic cartilage cDNA libraries
JOURNAL Osteoarthrits Cartilage 9 (7), 641-653 (2001)
MEDLINE 21482651
COMMENT Contact: Sanjay Kumar
          UW2109
          GlaxoSmithKline
          709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
          Tel: 610-270-7245
          Fax: 610-270-5598
          Email: sanjay_kumar-1@sk.com
          Seq primer: T7.
          Location/Qualifiers
          1..643
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_lib="HOA (Human Osteoarthritic Cartilage)"
          /tissue_type="cartilage"
          /lab_host="E.coli DH10 B"
          /note="vector: pSPORT 1; Site_1: Sali; Site_2: NotI;
          Directional"

BASE COUNT 197 a 119 c 158 g 169 t
ORIGIN

Query Match 42.6%; Score 368.6; DB 10; Length 643;
Best Local Similarity 79.9%; Pred. No. 1.1e-89;
Matches 434; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 20 atagatccaagacatgagactgtcagcaagaattatttggcttatatttgactgtttg 79
DB 95 AAAGATCCAAAAATGAGACTTCTAGCAAGATTTATTTGCCTTATCTTATGGCTATTTC 154

QY 80 tgtagcgaagattgtaaaggtcctctcccaagagaaaaattcagaattctctcaggttc 139
DB 155 TGTAGCAGAGATTGCAATGAACCTTCTCCAAGAAGAAATACAGAAATCTGACAGGTTTC 214

QY 140 gtggtctgaacaactatttcagaagcactcaggcaacctcacaattgccctctggata 199
DB 215 CTGCTCTGACCAACATATCTCAGAAAGGCCACCCAGGCTATCTATAATGCCGCCCTGGATA 274

QY 200 ccgacacttggtactattgttaaagtatgcagaatggagaatgggtgacctcttaaccc 259
DB 275 TAGATCTCTTGGAAATGTAATAATGTTATGCAGGAAGGGAGAAATGGGTGCTCTTAATCC 334

QY 260 atcaagatatgtcggaagggccatgtgggcatccccggagacacacccctttgggtccct 319
DB 335 ATTAAGGAAATGTGAGAAAGGCCCTGTGGACATCTCTGGAGATACTCTCTTTGGTACTTT 394

QY 320 taggctggcagttggatctgaattgattggttgcaaaaggtttttatcacatgtgatga 379
DB 395 TACCCTTACAGGAGGAAATGTGTGAATATGTTGTTAAAAGCTGTGTATACATTAATATGA 454

QY 380 agggtaaccaactattaggtgaaattgattaccgtgaatgtgatcagatgggtggaccaa 439
DB 455 GGGGTATCAATTGCTAGTCAGATTAATTACCGTGAATGTGACACATGGATGGACCAA 514

QY 440 tgatatccaatatgtgaagttgtgaagttgtgccagtgacagaactggagaatggaag 499

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2002, 19:15:26 ; Search time 537.93 Seconds
(without alignments)
2764.017 Million cell updates/sec

Title: US-09-316-163-2
Perfect score: 866
Sequence: 1 tcgagtcactctccaga.....ttgaattacatttgaaaaa 866

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	432.4	49.9	525	AAA43501	Mouse secreted exp
2	218.8	25.3	649	AAV02790	Human partial comp
3	217.2	25.1	581	AAV02791	Clone PR89PH410 C
4	46.2	5.3	3696	AAH75787	Receptor 222 codin
5	44.6	5.2	2539	AACT77331	Human ORFX ORF2886
6	44.6	5.2	2329	AAS03887	Human secreted pro
7	44.6	5.2	3804	ABA06548	Human cDNA SEQ ID
8	44.6	5.2	3804	AAS28968	cDNA encoding for
9	44.6	5.2	3804	AAS26868	Human cDNA encodin

ALIGNMENTS

RESULT 1

AAA43501 ID AAA43501 standard; cDNA; 525 BP.

XX AAA43501;

XX 21-AUG-2000 (first entry)

DE Mouse secreted expressed sequence tag SEQ ID NO:76.

Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
expressed sequence tag; EST; probe; chemotactic; proliferative;
immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
thrombolytic; antinflammatory; cytostatic; antibacterial; antifungal;
antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
antitumor; osteoprotective; neuroprotective; nootropic; antiproliferative;
cerebroprotective; anticonvulsant; antidepressant; gene therapy;
vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
central nervous system disorder; Alzheimer's disease; stroke;
Parkinson's disease; Huntington's disease; coagulation disorder;
haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
tumour; infection; depression; psoriasis; ss.

Mus musculus.

WO200021991-A1.

20-APR-2000.

15-OCT-1999; 99WO-US24206.

10	44.6	5.2	3804	22	AA531587	cDNA encoding nove
11	44.6	5.2	4088	22	AAK94919	Human full-length
12	44.6	5.2	5421	22	AAK94920	Human full-length
13	44.6	5.2	6153	24	ABA03880	Human POLY11 nucle
14	42.6	4.9	6131	24	ABL32890	Human immune syste
15	42.2	4.9	372	16	AAT23604	Human gene signatu
16	41.6	4.8	1430	22	AHH33539	Human colon cancer
17	41.6	4.8	2519	19	AAV54124	Human membrane pro
18	41.6	4.8	3139	21	AAC75833	Human ORFX ORF1388
19	41.6	4.8	14781	22	AAI36303	Human musculoskele
20	40	4.6	1878	22	AAI60621	Human polynucleoti
21	40	4.6	1927	22	AAI58835	Human polynucleoti
22	40	4.6	2404	22	AAH15794	Human cDNA sequenc
23	40	4.6	4590	22	AAH24065	Yeast AOD9604-asso
24	39.8	4.6	6131	24	ABL32891	Human immune syste
25	39.4	4.5	49999	20	AA223902	Human LOBO homolog
26	38.4	4.4	1640	17	AAT12162	Partial pJG4-5-CDK
27	37	4.3	6390	23	ABL12023	Drosophila melanog
28	37	4.3	12519	23	ABL12022	Drosophila melanog
29	36.8	4.2	1151	21	AAC40206	Arabidopsis thalia
30	36.8	4.2	2006	21	AAC42229	Arabidopsis thalia
31	36.2	4.2	385	18	AAV02797	Clone pZS3FH2576 #
32	36.2	4.2	385	18	AAV02798	Clone pZS3FH2576 #
33	36.2	4.2	771	22	AAK93472	Human cDNA clone r
34	36.2	4.2	14924	24	ABL32225	Human immune syste
35	35.6	4.1	1747	20	AAZ42234	Human normal bladd
36	35.6	4.1	1765	22	AAK94846	Human full-length
37	35.6	4.1	1880	22	AAH48345	Human granule memb
38	35.6	4.1	5409	21	AAA70151	Plasmodium falcipla
39	35.4	4.1	183	22	ABA74572	Human foetal liver
40	35.4	4.1	183	22	AAK23041	Human brain expres
41	35.4	4.1	183	22	AAK9218	Human bone marrow
42	35.4	4.1	183	22	AAI5053	Probe #23739 used
43	35.4	4.1	515	22	ABA62078	Human foetal liver
44	35.4	4.1	515	22	AAK10395	Human brain expres
45	35.4	4.1	515	22	AAK36298	Human bone marrow

QY 489 gagaatggaagaattgtgagtggtgcagccgaaaccagaccaggaataattatttttgacag 548
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 gagaatggaagaattgtcagtagtcaatggaaccagatcggaataaccattttggacaa 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 549 gtggtacgcttgaatgcaactccggttcaagattgaaggacagaaagaatgcaactgc 608
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 gcagtcaggttgtatgtatcaactcaggctacaagattgaaggagatgaagaatgcattgt 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 609 tcataaaatggcctctggagcaatgaaagccacagtggtgtgaaattttctgacctgcca 668
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 241 tcagacgatggttttggagttaaagagaaccaaagtgtgtgaaatttcattgcataatcc 300
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
QY 669 ccacgagttgaaatgga 686
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 301 ccagatgttataatgga 318
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
RESULT 3
AAV02791
ID AAV02791 standard; DNA; 581 BP.
XX AC AAV02791;
XX DT 27-APR-1998 (first entry)
XX DE Clone PRRB9FH410 CFH related protein DNA fragment.
XX KW Complement factor H; tumour associated antigen; renal cancer;
XX KW urogenital cancer; medicament; modulator; ss.
XX OS Synthetic.
XX XX WO9738136-A1.
XX PN 16-OCT-1997.
XX PD 09-APR-1997; 97WO-US05710.
XX PF 06-MAR-1997; 97US-0812481.
XX PR 09-APR-1996; 96US-0015083.
XX PR 09-APR-1996; 96US-0630048.
XX PR 06-MAR-1997; 97US-0038614.
XX XX (BARD-) BARD DIAGNOSTIC SCI INC.
XX PI Enfield DL, Hass GM, Kinders RJ;
XX DR WPI; 1997-512742/47.
XX DR P-PSDB; AAW39155.
XX PT Treating or screening for cancer, e.g. renal or urogenital cancer -
XX PT by modulating or detecting tumour associated human complement factor
XX PT H related antigen, or nucleic acid encoding it
XX PS Example 6B; Fig 6A; 104pp; English.
XX XX This partial cDNA sequence is present in clone PRRB9FH410 and encodes a
XX CC complement factor H related protein with homology to a region of the
XX CC human tumour-associated complement factor H (CFH). The detection of this
XX CC protein and a CFH antigen can be used in screening or for the treatment
XX CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
XX CC Agents that may modulate this antigen could be used in the manufacture of
XX CC a medicament for the treatment of a tumour cell.
XX XX Sequence 581 BP; 198 A; 85 C; 139 G; 159 T; 0 other;

Query Match 25.1%; Score 217.2; DB 18; Length 581;
Best Local Similarity 80.2%; Pred. No. 2.7e-52;
Matches 255; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 369 acatgtgatgaagggtaccactattagtgtaaatgattaccggtgaatgatgcagat 428

Db 1 acatgtgaatgaggggtatcaattgctagtgagatttaattaccgtgaatgtgacacagat 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 429 ggggtgaccaaataatgataatcattcaatgtgaagtgtgaagtgtccagtgacagaaactg 488
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ggaatgaccaaataatgataatcattcaatgtgaagtgtgaagtgttaccagtgacagaccca 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 489 gagaatggaagaattgtgagtggtgcagccgaaaccagaccaggaataattattttggacag 548
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 gagaatggaagaattgtcagtagtcaatggaaccagatcggaataaccattttggacaa 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 549 gtggtacgcttgaatgcaactccggttcaagattgaaggacagaaagaatgcaactgc 608
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 gcagtcaggttgtatgtatcaactcaggctacaagattgaaggagatgaagaatgcattgt 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 609 tcataaaatggcctctggagcaatgaaagccacagtggtgtgaaattttctgacctgcca 668
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 241 tcagacgatggttttggagttaaagagaaccaaagtgtgtgaaatttcattgcataatcc 300
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
QY 669 ccacgagttgaaatgga 686
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 301 ccagatgttataatgga 318
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
RESULT 4
AAH75787
ID AAH75787 standard; cDNA; 3696 BP.
XX AC AAH75787;
XX DT 16-OCT-2001 (first entry)
XX DE Receptor 222 coding sequence.
XX KW Receptor 222; cytostatic; anti-HIV; cancer; HIV infection; ss.
XX OS Unidentified.
XX FH Key Location/Qualifiers
XX FT CDS 402..3342
XX FT /*tag= a
XX FT /product= "Receptor 222"
XX CN1296952-A.
XX PN 30-MAY-2001.
XX PD 23-NOV-1999; 99CN-0124086.
XX PF 23-NOV-1999; 99CN-0124086.
XX PR (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.
XX PA Mao Y, Xie Y;
XX PI WPI; 2001-483894/53.
XX DR P-PSDB; AAG66398.
XX XX New polypeptide-complement receptor 222 for treating diseases, such as,
XX PT cancer and human immunodeficiency virus infection -
XX PS Claim 6; Pages 19-21 (Disclosure); 30pp; Chinese.
XX CC The present sequence is the coding sequence for receptor 222. Receptor
XX CC 222 and its coding sequence are useful in treating diseases e.g. cancer
XX CC and HIV infection.
XX SQ Sequence 3696 BP; 977 A; 837 C; 895 G; 987 T; 0 other;

Query Match 5.3%; Score 46.2; DB 22; Length 3696;
Best Local Similarity 59.5%; Pred. No. 0.0071;
Matches 78; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

RESULT	7	
ABA06548		
ID	ABA06548 standard; cDNA; 3804 BP.	
XX	AC	
XX	ABA06548;	
XX		
DT	10-JAN-2002 (first entry)	
XX		
XX	Human cDNA SEQ ID NO: 214.	
XX		
XX	Human; gene therapy; neural disorder; immune system disorder;	
KW	muscular disorder; reproductive disorder; gastrointestinal disorder;	
KW	pulmonary disorder; cardiovascular disorder; renal disorder;	
KW	proliferative disorder; inflammation; ss.	
XX		
XX	Homo sapiens.	
XX		
PN	WO200154474-A2.	
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001; 2001WO-US01349.	
XX		
PR	31-JAN-2000; 2000US-179065P.	
PR	04-FEB-2000; 2000US-180628P.	
PR	24-FEB-2000; 2000US-184664P.	
PR	02-MAR-2000; 2000US-186350P.	
PR	16-MAR-2000; 2000US-198874P.	
PR	17-MAR-2000; 2000US-190076P.	
PR	18-APR-2000; 2000US-198123P.	
PR	19-MAY-2000; 2000US-205515P.	
PR	07-JUN-2000; 2000US-209467P.	

PR	17-NOV-2000;	2000US-02492007
PR	17-NOV-2000;	2000US-0249208
PR	17-NOV-2000;	2000US-0249209
PR	17-NOV-2000;	2000US-0249210
PR	17-NOV-2000;	2000US-0249210
PR	17-NOV-2000;	2000US-0249211
PR	17-NOV-2000;	2000US-0249212
PR	17-NOV-2000;	2000US-0249213
PR	17-NOV-2000;	2000US-0249214
PR	17-NOV-2000;	2000US-0249215
PR	17-NOV-2000;	2000US-0249216
PR	17-NOV-2000;	2000US-0249217
PR	17-NOV-2000;	2000US-0249218
PR	17-NOV-2000;	2000US-0249244
PR	17-NOV-2000;	2000US-0249245
PR	17-NOV-2000;	2000US-0249264
PR	17-NOV-2000;	2000US-0249265
PR	17-NOV-2000;	2000US-0249297
PR	17-NOV-2000;	2000US-0249299
PR	17-NOV-2000;	2000US-0249300
PR	01-DEC-2000;	2000US-0250160
PR	01-DEC-2000;	2000US-0250391
PR	05-DEC-2000;	2000US-0251030
PR	05-DEC-2000;	2000US-0251988
PR	05-DEC-2000;	2000US-0256719
PR	06-DEC-2000;	2000US-0251479
PR	08-DEC-2000;	2000US-0251856
PR	08-DEC-2000;	2000US-0251868
PR	08-DEC-2000;	2000US-0251869
PR	08-DEC-2000;	2000US-0251989
PR	08-DEC-2000;	2000US-0251990
PR	11-DEC-2000;	2000US-0254097
PR	05-JAN-2001;	2001US-0259678
PA	(HUMA-) HUMAN GENOME SCI INCO	
XX		
PI	Rosen CA, Barash SC, Ruben	
XX		
PT	WPI; 2001-488777/53.	
DR	P-PSDB; AAU18126.	
DR		
XX		
PT	Isolated polypeptide and nucleic	
PT	acid encoding a protein capable of	
PT	preventing and/or prognosing	
PT	e.g. disorders associated wi	
XX		
PS	Claim 4; SEQ ID No 43; 524ppp	
XX		
CC	The present invention relates	
CC	to a method of identifying	
CC	uterine motility-association	
CC	and cDNA and genomic sequenc	
CC	of the sequences of the inventi	
CC	treatment, prevention and/or	
CC	with uterine motility such a	
CC	disorders. The polynucleotid	
CC	useful in gene therapy. AAS2	
CC	encoding for novel human ute	
CC	Note: The sequence data for	
CC	specification, but was obtain	
CC	at ftp.wipo.int/pub/publish	
XX		
SQ	Sequence 3804 BP; 1025 A; 85	
Query Match		
Best Local Similarity 5.2%;		
Matches 77; Conservative		
QY	531 gaattattattttggacagagggtg	
Db	1766 gattttgactgtgggaaggcgc	
QY	591 cagaagaataatgactgtctcata	
b	1926	

Sequence 3804 BP; 1025 A; 858 C; 915 G; 1006 T; 0 other

Qy	531	gaatattatttggacaggtgtagctgttgaatgcaactcggcttcaagattgaagga	590
Db	1766	gatttgactgtgaaagcgacgcggatcagtgctcaaggcttcaagctctagga	1825
Qy	591	cagaaagaaatgcactgctcataaataggctctggagcaatgaaagccacagtggtg	650
Db	1826	cttctgaaatccactgtgaagcgcgatggcagtgagctctggtgtccccactgtgaa	1885

OY 651 gaaattcttg 661
| | | | |
Db 1886 cacactcttg 1896

RESULT 9
ID AAS26868
AC AAS26868 standard; cDNA; 3804 BP.
XX AAS26868;
XX
XX
XX 07-NOV-2001 (first entry)
DT Human cDNA encoding a novel secreted protein, SEQ ID 60.
DE XX
DE Human; immunosuppressive; antiarthritic; ss; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
XX
PN WO20015441-A2.
PD
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01320.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249220.
 PR 17-NOV-2000; 2000US-0249221.
 PR 17-NOV-2000; 2000US-0249222.
 PR 17-NOV-2000; 2000US-0249223.
 PR 17-NOV-2000; 2000US-0249224.
 PR 17-NOV-2000; 2000US-0249225.
 PR 17-NOV-2000; 2000US-0249226.
 PR 17-NOV-2000; 2000US-0249227.
 PR 17-NOV-2000; 2000US-0249228.
 PR 17-NOV-2000; 2000US-0249229.
 PR 17-NOV-2000; 2000US-0249230.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PR (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI: 2001-476222/51.

XX P-PSDB; AAU16963.

XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
 PT diagnose diseases or disorders associated with aberrant expression or
 PT activity of polypeptides, for treating blood clotting disorder,
 PT haemophilia

XX Claim 1; SEQ ID NO 60; 601pp: English.

XX The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiodysplasia, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present

Query Match 5.2%; Score 44.6; DB 22; Length 3804;

Best Local Similarity 58.8%; Pred. No. 0.021;

Matches 77; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 531 gaattatttttgacagtggtgacgtcttgatgaactcccgcttcaagattgaagga 590

Db 1766 gattttgactgtggaaggcagcccgaggtattcagttcaaggctccctcaggga 1825

QY 591 cagaaagaatgcactgtctcataaaatggcctctggagcaatgaaagccacagtggtg 650

Db 1826 cttctgaaatcacctgtggaagccgagtgccagtgagctctgggttcccccaactgtgaa 1885

QY 651 gaaatttcttg 661
 | | | | | | | | | |

Db 1886 cacacttcttg 1896

RESULT 10
 AAS31587
 ID AAS31587 standard; cDNA; 3804 BP.
 XX
 AC AAS31587;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE cDNA encoding novel human calcium-binding protein #11.
 XX
 KW Human; calcium-binding protein; calcium flux; neurological disease;
 KW immune dysfunction; digestive disorder; neoplastic disease;
 KW blood disorder; infectious disease; gene therapy; immunosuppressive;
 KW antiarthritic; cytostatic; vasotropic; antibacterial; nootropic;
 KW virucide; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200155304-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01302.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0217496.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 23-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.

```
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 23-SEP-2000; 2000US-0234977.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
```

```
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
```

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465568/50.

P-PSDB; AAU19902.

Isolated nucleic acid molecule encoding a calcium-binding protein is used in preventing, treating or ameliorating a medical condition -

Claim 4; SEQ ID No 21; 542pp; English.

The present invention relates to the isolation of novel human calcium-binding proteins (AAU1982-AAU19969), and cDNA and genomic sequences encoding for these proteins. The sequences of the invention are useful in the diagnosis, prevention and/or prognosis of diseases associated with aberrant calcium flux. Such disorders include neurological diseases (e.g. amyotrophic lateral sclerosis, ALS), immune dysfunction (e.g. severe combined immunodeficiency, SCID), digestive disorders (e.g. irritable bowel syndrome, IBS), neoplastic disease (e.g. cancer), blood disorders (e.g. haemophilia), and/or infectious disease (e.g. acquired immunodeficiency syndrome, AIDS). The novel calcium-binding proteins are also useful as screening tools to identify antagonists and/or agonists that may enhance or inhibit activities mediated by calcium-binding proteins. The polynucleotides of the invention are also useful in gene therapy. AAS31577-AAS31634 represent cDNA sequences encoding for the novel human calcium-binding proteins.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 3804 BP; 1025 A; 858 C; 915 G; 1006 T; 0 other;

Query Match 5.2%; Score 44.6; DB 22; Length 3804;
Best Local Similarity 58.88; Pred. No. 0.021;
Matches 77; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 531 gaatattattttggacaggttggtacgtttgaatgcactccggcttcaagattgaagga 590

Db 1766 gatttgactgtggaagggcagccggatttcaggttcaagagcttcaagctcttaga 1825

QY 591 cagaaagaatgcactgtctcataaaatggcctctggagcaatgaaaaagccacagtgtgtg 650

Db 1826 ctttctgaaatcacctgtggaagccgatggccagtggtgagctctgtggttcccccctgtgaa 1885

QY 651 gaaatttcttg 661

Db 1886 cacacttcttg 1896

RESULT 11

AAK94919
ID AAK94919 standard; cDNA; 4088 BP.

XX AC
XX DT
XX AC
XX AAK94919;

DT 06-NOV-2001 (first entry)

XX Human full-length cDNA, SEQ ID NO: 4150.

DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX KW
XX OS
XX Homo sapiens.

XX PN
XX EP1130094-A2.

XX PD
XX 05-SEP-2001.

XX PF
XX 07-JUL-2000; 2000EP-0114089.

XX PR
XX 08-JUL-1999; 99JP-0194486.

XX PR
XX 11-JAN-2000; 2000JP-0118774.

XX PR
XX 02-MAY-2000; 2000JP-0183765.

XX PA
(HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX DR
XX WPI; 2001-524255/58.

XX DR
XX P-PSDB; AAM93953.

XX PT
XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -

XX PS
XX Claim 8; SEQ ID NO 4150; 1380pp + sequence listing; English.

XX CC
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a full length
XX human cDNA of the invention.

XX CC
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.

XX SQ
XX Sequence 4088 BP; 1067 A; 939 C; 979 G; 1103 T; 0 other;

Query Match 5.2%; Score 44.6; DB 22; Length 4088;
Best Local Similarity 58.8%; Pred. No. 0.021;
Matches 77; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 531 gaattattttggacaggtgtacgttgaatgcaactcggcttcaagattgaagga 590

DB 2076 gattttgactgtggaaggcagccggttcagttcctcaaggcttcaagctcctagga 2135

OY 591 cagaaagaatgcactgtctcataaaatggcctcttgagcaatgaaagccacagtggtg 650

DB 2136 cttctgaaatcacctggaagccgagtgagcagtggtggttccccccactgtgaa 2195

OY 651 gaaattttcttg 661

DB 2196 cacactttcttg 2206

RESULT 12

AAK94920

ID AAK94920 standard; cDNA; 5421 BP.

XX AC
XX AAK94920;

XX DT
XX 06-NOV-2001 (first entry)

XX DE
XX Human full-length cDNA, SEQ ID NO: 4152.

XX KW
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX OS
XX Homo sapiens.

XX PN
XX EP1130094-A2.

XX PD
XX 05-SEP-2001.

XX PF
XX 07-JUL-2000; 2000EP-0114089.

XX PR
XX 08-JUL-1999; 99JP-0194486.

XX PR
XX 11-JAN-2000; 2000JP-0118774.

XX PR
XX 02-MAY-2000; 2000JP-0183765.

XX PA
(HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX DR
XX WPI; 2001-524255/58.

XX DR
XX P-PSDB; AAM93954.

XX PT
XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -

XX PS
XX Claim 8; SEQ ID NO 4152; 1380pp + sequence listing; English.

XX CC
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a full length
XX human cDNA of the invention.

XX CC
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.

XX SQ
XX Sequence 5421 BP; 1398 A; 1275 C; 1318 G; 1430 T; 0 other;

Query Match 5.2%; Score 44.6; DB 22; Length 5421;
Best Local Similarity 58.8%; Pred. No. 0.024;
Matches 77; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 531 gaattattttggacaggtgtacgttgaatgcaactcggcttcaagattgaagga 590

DB 3515 gattttgactgtggaaggcagccggttcagttcctcaaggcttcaagctcctagga 3574

OY 591 cagaaagaatgcactgtctcataaaatggcctcttgagcaatgaaagccacagtggtg 650

DB 3575 cttctgaaatcacctggaagccgagtgagcagtggttccccccactgtgaa 3634

OY 651 gaaattttcttg 661

DB 3635 cacactttcttg 3645

RESULT 13

ABA03880

ID ABA03880 standard; cDNA; 6153 BP.

XX AC

ABA03880;

XX	14-FEB-2002	(first entry)
DT		
XX		
DE	Human POLY11	nucleotide sequence SEQ ID NO:21.
XX		
KW	Human; POLYX; gamma aminobutyric acid receptor; GABA receptor;	
KW	epidermal growth factor; EGF; complement receptor; HSPC; syntaxin;	
KW	haematopoietic stem and progenitor cell; sulphotransferase; prohibitin;	
KW	antidepressant; cerebroprotective; antiparkinsonian; nootropic; relaxant;	
KW	anticonvulsant; neuroleptic; neuroprotective; antialcoholic; cardiant;	
KW	tranquilliser; antiarrhythmic; psychiatric; medical; depression; stroke;	
KW	Parkinson's disease; Huntington's disease; Tourette's syndrome; anxiety;	
KW	amyotrophic lateral sclerosis; head trauma; Alzheimer's disease;	
KW	alcoholism; vigilance; muscle tension; epileptogenic; memory function;	
KW	cardiomyopathy; arrhythmogenic right ventricular dysplasia; ss.	
XX		
OS	Homo sapiens.	
XX		
EH	Key	Location/Qualifiers
FT	CDS	1..6153
FT		/*tag= a
FT		/product= "POLY11"
FT		/transl_except= (pos:1099..1101,aa:Xaa)
FT		/note= "Xaa is unknown"
XX		
PN	WO200179294-A2.	
XX		
PD	25-OCT-2001.	
XX		
PF	19-APR-2001; 2001WO-US12854.	
XX		
PR	19-APR-2000; 2000US-198293P.	
PR	20-APR-2000; 2000US-198645P.	
PR	25-APR-2000; 2000US-199476P.	
PR	26-APR-2000; 2000US-199880P.	
PR	26-APR-2000; 2000US-200024P.	
PR	26-APR-2000; 2000US-200025P.	
PR	09-JUN-2000; 2000US-210809P.	
PR	17-JUL-2000; 2000US-218591P.	
PR	11-AUG-2000; 2000US-224610P.	
PR	09-FEB-2001; 2001US-267673P.	
XX		
PA	(CURA-) CURAGEN CORP.	
XX		
PI	Taupter RJ, Vernet CM, Fernandes E, Shimkets RA, Majumder K;	
PI	Padigaru M, Colman SD, Zerhusen BD, Spytek KA, Burgess CE, Liu X;	
XX		
XX	WPI: 2002-017601/02.	
DR	P-PSDB; AAG68264.	
DR		
XX		
PT	New isolated polypeptides for treating a broad range of pathological	
PT	states, e.g., depression, stroke, Parkinson's disease, Huntington's	
PT	disease, Tourette's syndrome, amyotrophic lateral sclerosis, head	
PT	trauma, and Alzheimer's	
XX		
PS	Claim 9; Page 43-45; 155pp; English.	
XX		
CC	The present invention describes polypeptides (I), designated POLYX	
CC	polypeptides (e.g. POLY1, POLY2, etc.), and the polynucleotide	
CC	sequences (II) encoding them. POLY1-4 are members of the gamma	
CC	aminobutyric acid (GABA) receptor family; POLY5-8 are members of the	
CC	epidermal growth factor (EGF) family; POLY9-11 are members of the	
CC	complement receptor family; POLY12 is a member of the haematopoietic	
CC	stem and progenitor cell (HSPC) family; POLY13 is a member of the	
CC	sulphotransferase family; POLY14-16 are members of the syntaxin family;	
CC	and POLY17 is a member of the prohibitin family. (I) and (II) can have	
CC	antidepressant, cerebroprotective, antiparkinsonian, nootropic, relaxant,	
CC	anticonvulsant, neuroleptic, neuroprotective, antialcoholic, cardiant,	
CC	tranquilliser and antiarrhythmic activities. (I) and (II) can be used	
CC	for treating or preventing a POLYX-associated disorder in humans as a	
CC	therapeutic in the manufacture of a medicament for treating a syndrome	
CC	associated with a human disease selected from a POLYX-associated	
CC	disorder, for treating a pathological state in a mammal, especially	

patients suffering from, e.g., psychiatric and medical conditions, depression, stroke, Parkinson's disease, Huntington's disease, Tourette's syndrome, ankyrotrophic lateral sclerosis, head trauma, Alzheimer's disease, alcoholism, vigilance, anxiety, muscle tension, epileptogenic activity and memory functions, cardiomyopathy and arrhythmogenic right ventricular dysplasia. The present sequence encodes POLY11.

Sequence 6153 BP; 1610 A; 1482 C; 1563 G; 1497 T; 1 other;

Query Match / 5.2%; Score 44.6; DB 24; Length 6153;
Best Local Similarity 58.8%; Pred. No. 0.025;
Matches 77; Conservative 0; Mismatches 54; Indels 0; Gaps 0

Qy 531 gaataattatttgacacagtgtagcgtttgaataccaactcggcttcaagattgaagga 590
Db 3280 gatttgactgtggaagcgccggttcagtgcttcaaggcttcaagctctaggga 3339
Qy 591 cagaagaaatgcactgctcataaaatggcctctggagcaatgaaaagccacagtgtgtg 650
Db 3340 cttctgaaatcaactgtgaagcgcgatggcagtgaggctctgggttcccccactgtgaa 3399
Qy 651 gaaattttcttg 661
Db 3400 cacacttttg 3410

RESULT 14
ABL32890
ID ABL32890 standard; DNA; 6131 BP.
AC ABL32890;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 863.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
XX OS
XX WO200200928-A2.
XX PN
XX 03-JAN-2002.
XX PD
XX 02-JUL-2001; 2001WO-EP07537.
XX PF
XX 30-JUN-2000; 2000DE-1032529.
XX PR
XX 01-SEP-2000; 2000DE-1043826.
XX XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation
XX
XX
XX Claim 1; SEQ ID NO 863; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system diseases.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:00:49 ; Search time 39.15 Seconds
(without alignments)
508.059 Million cell updates/sec

Title: US-09-316-163-9
Perfect score: 1163
Sequence: 1 EDGNELPPRRNTILGTSM.....VEISCKSPDVINGSPISQKI 207

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1163	100.0	449	1 NBHUHS	complement factor
2	1163	100.0	1231	1 NBHUH	complement factor
3	822	70.7	1234	1 NBMSH	complement factor
4	499	42.9	669	2 S6551	factor H - bovine
5	368.5	31.7	1053	2 S46199	probable complement
6	275	23.6	360	2 T42921	complement control
7	262	22.5	597	1 S53711	C4BP alpha chain p
8	261	22.4	381	1 B26359	decay-accelerating
9	261	22.4	440	2 A26359	decay-accelerating
10	260	22.4	676	2 A45900	complement C3b rec
11	255	21.9	2489	2 J73012	E-selectin - pig
12	253.5	21.8	482	2 JC5092	decay-accelerating
13	251	21.6	340	2 I56234	complement recepto
14	247	21.2	2014	2 I36936	apolipoprotein H h
15	246	21.2	263	1 WNVZSP	C4b-binding protei
16	244	21.0	597	1 NBHU4	complement regulat
17	243.5	20.9	497	2 JC2054	hypothetical prote
18	243	20.9	560	2 T16833	complement recepto
19	240.5	20.7	440	2 A43519	sperm CD46 - human
20	238.5	20.5	349	2 G02913	membrane cofactor
21	238.5	20.5	369	2 I57998	membrane cofactor
22	238.5	20.5	377	2 I54479	membrane cofactor
23	238.5	20.5	384	2 S01896	membrane cofactor
24	236	20.3	263	1 C36838	complement control
25	236	20.3	263	2 T28450	hypothetical prote
26	236	20.3	302	1 WMBE1E	secretory compleme
27	236	20.3	360	1 WMBR2E	membrane-bound com
28	235	20.2	263	2 B72152	B18L protein - var
29	235	20.2	482	2 A34924	complement C3b/C4b

30 230.5 19.8 610 2 A35046 E-selectin precurs
31 229.5 19.7 362 2 JC5194 membrane cofactor
32 229.5 19.7 369 2 JC5138 membrane cofactor
33 228 19.6 469 1 NBMSC4 C4b-binding protei
34 227.5 19.6 768 2 A42755 p-selectin precurs
35 227.5 19.6 830 2 A30359 p-selectin precurs
36 222 19.1 768 2 I53821 p-selectin - rat
37 218.5 18.8 612 2 B42755 E-selectin precurs
38 216 18.6 558 2 S57953 C4BP protein alpha
39 216 18.6 977 2 I52657 seizure-related pr
40 214.5 18.4 551 2 I46709 endothelial leukoc
41 214.5 18.4 610 1 I46001 C4b-binding protei
42 212 18.2 1025 1 A43526 complement C3d/Eps
43 210.5 18.1 646 2 JN0473 p-selectin precurs
44 210.5 18.1 1091 1 PL0009 complement C3d/Eps
45 210 18.1 372 1 A32375 L-selectin precurs

ALIGNMENTS

RESULT 1
NBHUHS
N:Alternate names: complement factor H-related protein; complement protein H
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 23-Feb-1996 #text_change 08-Dec-2000
C:Accession: S03013; B60238; A27877; A61103; A26505; S10479
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A:Title: The complete amino acid sequence of human complement factor H.
A:Reference number: S00254; MUID:88134059
A:Accession: S03013
A:Molecule type: mRNA
A:Residues: 1-449 <RIP>
A:Cross-references: EMBL:X07523; EMBL:Y00716; NID:932492; PIDN:CAA30403.1; PID:975807
A:Note: part of this sequence, including the amino end of the mature protein was conf
A:Note: 402-Tyr was also found
R:Estaller, C.; Schwaebler, W.; Dierich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A:Title: Human complement factor H: two factor H proteins are derived from alternativ
A:Reference number: A60238; MUID:91184292
A:Accession: B60238
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-33;434-449 <EST>
A:Note: Only portions of this 1.8 kilobase mRNA were sequenced
R:Schulz, T.F.; Schwaebler, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.
Eur. J. Immunol. 16, 1351-1355, 1986
A:Title: Human complement factor H: isolation of cDNA clones and partial cDNA sequenc
A:Reference number: A27877; MUID:87054207
A:Accession: A27877
A:Molecule type: mRNA
A:Residues: 11',55-401,'Y',403-449 <SCH>
A:Cross-references: GB:X04697; NID:931991; PIDN:CAB41739.1; PID:94725976
A:Note: an additional nucleotide present within the codon for Glu-310 was thought to
R:Schwaebler, W.; Zwirner, J.; Schulz, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.
Eur. J. Immunol. 17, 1485-1489, 1987
A:Title: Human complement factor H: expression of an additional truncated gene produc
A:Reference number: A61103; MUID:88055295
A:Accession: A61103
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 27-76 <SC2>
A:Note: this is a partial sequence of an alternatively spliced 1.8 kilobase mRNA that
R:Sim, R.B.; DiScipio, R.G.
Biochem. J. 205, 285-293, 1982
A:Title: Purification and structural studies on the complement-system control protein
A:Reference number: A26505; MUID:83048213
A:Accession: A26505
A:Molecule type: protein
A:Residues: 19-20,'O',22-29,'V',31-33,'O',35 <SIM>
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P

Biochemistry 31, 3626-3634, 1992
A:Title: Solution structure of the fifth repeat of factor H: A second example of the complement factor H
A:Reference number: A44551; MUID:92232649
A:Contents: annotation; NMR structure determination, residues 264-292
R:Kristensen, T.; Wetzel, R.A.; Tack, B.F.
J. Immunol. 136, 3407-3411, 1986
A:Title: Structural analysis of human complement protein H: homology with C4b binding protein
A:Reference number: S10479; MUID:86169701
A:Accession: S10479
A:Molecule type: mRNA
A:Residues: 226-401, 'Y', 403-449 <KRI>
A:Cross-references: GB:M12303; NID:g180472; PIDN:AAA52013.1; PID:g180473
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. Homologous transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.
C:Genetics: <HF1>
A:Gene: GDB:HF1; HF
A:Cross-references: GDB:120041; OMIM:134370
A:Map position: lq32-lq32
C:Genetics: <HF2>
A:Gene: GDB:HF2; HF
A:Cross-references: GDB:129095
A:Map position: lq32-lq32
A:Note: the correspondence between the two loci and the sequences indicated is unclear;
C:Function:
A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increases the alternative complement pathway
A:Pathway: complement alternate pathway
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-449/Product: complement factor H, short splice form #status experimental <MAT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:21-66, 52-80, 85-129, 114-141, 146-192, 178-205, 210-251, 237-262, 267-309, 294-320, 325-374, 357-374/Binding site: carbohydrate (Asn) (covalent) #status absent
F:217/Binding site: carbohydrate (Asn) (covalent) #status absent
Query Match 100.0%; Score 1163; DB 1; Length 449;
Best Local Similarity 100.0%; Pred. No. 1.9e-90;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EDCNELPRNTEILTGSDQTYEGTQAIYKCRPGYRSLGNVIMVCRKGHWALNPLR 60
DB 19 EDCNELPRNTEILTGSDQTYEGTQAIYKCRPGYRSLGNVIMVCRKGHWALNPLR 78
QY 61 KCKRPGCHPGDTPGTFTLTGNNVFYGVKAVYTCNKGQYLLGEINRECDTDCGWTNDI 120
DB 79 KCKRPGCHPGDTPGTFTLTGNNVFYGVKAVYTCNKGQYLLGEINRECDTDCGWTNDI 138
QY 121 PICEVVKCLPVTAPENKIVSAMDPREYHFGQAVRVCNNGYKIEGDEMHCSDDGFW 180
DB 139 PICEVVKCLPVTAPENKIVSAMDPREYHFGQAVRVCNNGYKIEGDEMHCSDDGFW 198
QY 181 SKEKPKCIVEISCKSPDVINGSPISOKI 207
DB 199 SKEKPKCIVEISCKSPDVINGSPISOKI 225
RESULT 2
NBHUH
complement factor H precursor, long splice form [validated] - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Dec-2000
C:Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A:Title: The complete amino acid sequence of human complement factor H.

A:Reference number: S00254; MUID:88134059
A:Accession: S00254
A:Molecule type: mRNA
A:Residues: 1-1231 <RIP>
A:Cross-references: EMBL:Y00716; NID:g31964; PIDN:CAA68704.1; PID:g31965
A:Note: 402-Tyr was also found
A:Note: parts of this sequence, including the amino and carboxyl ends of the mature protein, are identical to those of factor H.
R:Estaller, C.; Schwaebler, W.; Dierich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A:Title: Human complement factor H: two factor H proteins are derived from alternative splicing of a single gene
A:Reference number: A60238; MUID:91184292
A:Accession: A60238
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-56; 117-1231 <EST>
A:Note: only portions of this 4.3 kilobase mRNA were sequenced
R:Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.
Biosci. Rep. 7, 201-207, 1987
A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human complement factor H
A:Reference number: A54726; MUID:88025472
A:Accession: A54726
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'DFRN', 579-1231 <DAY>
A:Cross-references: GB:M17517; NID:g180497; PIDN:AAA52016.1; PID:g180498
A:Note: parts of this sequence were determined by protein sequencing
R:Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.
Biosci. Rep. 6, 65-72, 1986
A:Title: Partial characterization of human complement factor H by protein and cDNA sequencing
A:Reference number: A61565; MUID:86188123
A:Accession: A61565
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'METGRNHLNAKI', 1050-1057, 'T', 1059-1102 <RI2>
R:Sim, R.B.; DiScipio, R.G.
Biochem. J. 205, 285-293, 1982
A:Title: Purification and structural studies on the complement-system control protein factor H
A:Reference number: A26505; MUID:83048213
A:Accession: A26505
A:Molecule type: protein
A:Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P.
Biochemistry 31, 3626-3634, 1992
A:Title: Solution structure of the fifth repeat of factor H: A second example of the complement factor H
A:Reference number: A44551; MUID:92232649
A:Contents: annotation; NMR structure determination, residues 264-292
R:Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.
J. Mol. Biol. 219, 717-725, 1991
A:Title: Three-dimensional structure of a complement control protein module in solution
A:Reference number: A49224; MUID:91278097
A:Contents: annotation; NMR structure determination, residues 927-985
R:Estaller, C.; Koistinen, V.; Schwaebler, W.; Dierich, M.P.; Weiss, E.H.
J. Immunol. 146, 3190-3196, 1991
A:Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel gene
A:Reference number: I56100; MUID:91201892
A:Accession: I72654
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1047-1231 <RES>
A:Cross-references: GB:M65294; NID:g183766; PIDN:AAA35948.1; PID:g183767
R:Carroll, J.A.; Bates, R.C.; Smith, A.I.; Tetoz, T.; Arellano, A.; Gordon, D.L.; Burn
Biochim. Biophys. Acta 1289, 305-311, 1996
A:Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.
A:Reference number: S66298; MUID:96205365
A:Accession: S66298
A:Status: preliminary
A:Molecule type: protein
A:Residues: 411-419; 574-580-582 <CAR>
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner.
C:Genetics: <HF1>
A:Gene: GDB:HF1; HF

Db 139 PLCEVYKCLPVTLENGRIVSGAAETDQYFQGVVRFECNSGFKTEGHKEIHCSENGLW 198
QY 181 SREKPKVCEISKSPDVINGSPISOK 206
Db 199 SNEKPRCVELICTPPRVENGDDGINKV 224

RESULT 4

S65551
Query Match 42.9%; Score 499; DB 2; Length 669;
Best Local Similarity 64.9%; Pred. No. 2.5e-34;
Matches 85; Conservative 19; Mismatches 27; Indels 0; Gaps 0;
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S65551
R:Soames, C.J.; Day, A.J.; Sim, R.B.
Biochem. J. 315, 523-531, 1996
A:Title: Prediction from sequence comparisons of residues of factor H involved in the in
A:Reference number: S65551; MUID: 96202005
A:Accession: S65551
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-669 <SOA>
A:Cross-references: GB:X98697; NID:q1419423; PIDN:CAA67257.1; PID:q1419424
C:Superfamily: complement factor H; complement factor H repeat homology
F:55-114/Domain: complement factor H repeat homology <FHR1>
F:296-349/Domain: complement factor H repeat homology <FHR2>
F:355-412/Domain: complement factor H repeat homology <FHR3>
F:416-471/Domain: complement factor H repeat homology <FHR4>
F:476-530/Domain: complement factor H repeat homology <FHR>
F:538-592/Domain: complement factor H repeat homology <FHR5>
F:599-651/Domain: complement factor H repeat homology <FHR6>

QY 76 GFTLTGGVFEYGVKAVYTCNEGYYQLLGEINRECDTGTGWTNDIPICEVVKCLPVTAPPE 135
Db 3 GSPHLAEGNOFEYGAIVYTCDEGYQVMGEMNFREDTGTGWTNDIPICEVVKCLPVTPEE 62
QY 136 NGKIVSAMPEPREYHFGQAVFVCSNGYKIEGDEMHCSDDGFSWKEKPKVCEISKSP 195
Db 63 NCKIFSDALEPQDEYTYGQVVOFCNSGYMLDGPQKHCSAGVWSAETPKVCEIFCKPP 122
QY 196 DVINGSPISOK 206
Db 123 VILNGQAVLPK 133

RESULT 5

S46199
Query Match 42.9%; Score 499; DB 2; Length 669;
Best Local Similarity 64.9%; Pred. No. 2.5e-34;
Matches 85; Conservative 19; Mismatches 27; Indels 0; Gaps 0;
C:Species: Parabrax nebulifer
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 18-Jul-2001
C:Accession: S46199; S77894
R:Dahmen, A.; Kaidoh, T.; Zipfel, P.F.; Gagli, I.
Biochem. J. 301, 391-397, 1994
A:Title: Cloning and characterization of a cDNA representing a putative complement-regul
A:Reference number: S46199; MUID: 94318039
A:Accession: S46199
A:Molecule type: mRNA
A:Residues: 1-1053 <DAH1>
A:Cross-references: EMBL:L21703; NID:g639894; PIDN:AAA92556.1; PID:g639895
A:Experimental source: liver
A:Keywords: S77894
A:Molecule type: protein
A:Residues: 526-532, 'X', 534-537, 809-817, 'X', 819-826 <DAH2>
C:Genetics:
A:Gene: SBI
C:Superfamily: complement factor H repeat homology
C:Keywords: glycoprotein
F:89-145/Domain: complement factor H repeat homology <FH01>

F:334-389/Domain: complement factor H repeat homology <FH02>
F:450-502/Domain: complement factor H repeat homology <FHR1>
F:569-624/Domain: complement factor H repeat homology <FH03>
F:682-738/Domain: complement factor H repeat homology <FH04>
F:743-802/Domain: complement factor H repeat homology <FHR>
F:935-989/Domain: complement factor H repeat homology <FH05>
F:993-1052/Domain: complement factor H repeat homology <FH06>

Query Match 31.7%; Score 368.5; DB 2; Length 1053;
Best Local Similarity 39.6%; Pred. No. 3.9e-23;
Matches 72; Conservative 25; Mismatches 76; Indels 9; Gaps 4;

QY 21 DQTYPEGTQAIYKCRPGYSLGNVIMVCRKGEWVALNPLRKCQKRCRPGHGDPTFGFTTL 80
Db 47 EASYPGGRQVRVCCNVGYS--GFFKLVCVGGKWETRG--AKQPRSCGHPGDAQADFHL 102
QY 81 TGNVFEYGVKAVYTCNEGYYQLLGEINRECDTGTGWTNDIPICEVVKCLPVTAPENGKIV 140
Db 103 AEGNDFVFGSKVYTCQKQYMSRINRYRCVAEGWDGVVPVCGESQCCPLIHVDNNVQV 162
QY 141 SSAMEPDRVHFQAVRFVCSNGYKI--EGDEMHCSDDGFSWKEKPKVCEISKSPDVIN 199
Db 163 GG-----PEATPGNVVRFSCSRSEILDGSPEDYCDERGDWSGVPVCKAITCAIPPIEN 218
QY 200 GS 201
Db 219 GN 220

RESULT 6

T42921
Query Match 31.7%; Score 368.5; DB 2; Length 1053;
Best Local Similarity 39.6%; Pred. No. 3.9e-23;
Matches 72; Conservative 25; Mismatches 76; Indels 9; Gaps 4;
C:Species: ateline herpesvirus 3
A:Variety: strain 73
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42921
R:Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A:Description: Primary structure of the herpesvirus ateles genome.
A:Reference number: 222274
A:Accession: T42921
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-360 <ALB>
A:Cross-references: EMBL:AF083424; PIDN:AAC95530.1
A:Experimental source: strain 73
C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom

Query Match 23.6%; Score 275; DB 2; Length 360;
Best Local Similarity 37.1%; Pred. No. 9.4e-16;
Matches 78; Conservative 20; Mismatches 80; Indels 32; Gaps 14;

QY 8 PRNTEI-----LTGSDSDQTYPEGTQAIYKCRPGYSLGNVI--MVCRKGEWVALNPL 59
Db 24 PKRNVYSLRYVNITN--SSGYPNGTTLQVTCRKGY--IGRQIQVTCVNGNWTVPN-- 77
QY 60 RKCQKRCPCGHPGTPGTFTLTGGNVFEYGVKAVYTCNEGYYQLLGEINRECDTGTGWTND 115
Db 78 -ECQKRCSTPADLLNGWTVT--GNLY--YGSVITYTCNTGYQLLGSPT--SSCLLGGPDGRV 133
QY 116 -WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRVHFQAVRFVCSNGYKI--EGDEMHCSDDG 171
Db 134 NMTPRPPEITCKRPPPTIANGTHTNI-----KEYTYLDAVYTCNDETTLTLTGPS 188
QY 172 MHCSDDGFV--SKEKPKVCEISKSPDVING 200
Db 189 KQCSSETGRVWPDEETKCEKVCCKIPQVANG 218

RESULT 7

S53711

C4BP alpha chain precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S53711
R:de Frutos, P.G.; Dahlbaeck, B.
Biochim. Biophys. Acta 1261, 285-289, 1995
A:Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence m
A:Reference number: S53711; MUID:95226458
A:Accession: S53711
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-597 <DEF>
A:Cross-references: EMBL:Z35490
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F:50-107/Domain: complement factor H repeat homology <FH1>
F:112-169/Domain: complement factor H repeat homology <FH2>
F:174-234/Domain: complement factor H repeat homology <FH3>
F:239-294/Domain: complement factor H repeat homology <FH4>
F:299-360/Domain: complement factor H repeat homology <FH5>
F:364-422/Domain: complement factor H repeat homology <FH6>
F:426-480/Domain: complement factor H repeat homology <FH7>
F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match 22.5%; Score 262; DB 1; Length 597;
Best Local Similarity 32.2%; Pred. No. 2e-14; Mismatches 93; Indels 22; Gaps 10;
Matches 69; Conservative 30

QY 2 DCNELPPRRNTEILTGSWSDDQTYPECTQAIYKCRPGYSLG-NVIMVCR-KGEWVALNPL 59
DB 49 DCGP-PPHLLFASSISELSENEVQTGILKTCRPGYTRNGLNPILTKRPLGLW---SYD 104
QY 60 RKCQRPCHGPDTPGTTFLTGGNVFEGVAVYTCNGYQLLGEINRECDT-----G 115
DB 105 TFCVKRRCRNPGLPNGQVEK--TDFSGQSFSESGEYILIGSTT-SHCDIQEKGEV 161
QY 116 WTNDIPICVWVKCLPVTAPENKIVSSAMEPDREXH-FGOAVRFVCSNGYKIEGDEEMHC 174
DB 162 WSDLPKKEIVACEPPNLIINGKHGG-----NEDHTYGSSTYSCNPRFSLLEASISC 217
QY 175 S-----DDGFWMSKEKPKCBEISKSPDVINGSPIS 204
DB 218 TVKNKTVGVWSPSPVCKEIIICSPNPVPHGKIIS 251

RESULT 8
B26359
decay-accelerating factor, GPI-anchored splice form precursor - human
N:Alternate names: CD55; DAF
C:Species: Homo sapiens (man)
C:Date: 05-Oct-1988 #sequence_revision 16-Aug-1996 #text_change 19-Jan-2001
C:Accession: B26359; A27666; A39101; I52594; I52564
R:Caras, I.W.; Davitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.
Nature 325, 545-549, 1987
A:Title: Cloning of decay-accelerating factor suggests novel use of splicing to generate
A:Reference number: A26359; MUID:87115645
A:Accession: B26359
A:Molecule type: mRNA
A:Residues: 1-381 <CAR>
A:Cross-references: GB:M30142; NID:g181464; PIDN:AAA52168.1; PID:g181465
R:Medof, M.E.; Lublin, D.M.; Hollers, V.M.; Ayers, D.J.; Getty, R.R.; Leykam, J.F.; Atkin
Proc. Natl. Acad. Sci. U.S.A. 84, 2007-2011, 1987
A:Title: Cloning and characterization of cDNAs encoding the complete sequence of decay-a
A:Reference number: A27666; MUID:87175602
A:Accession: A27666
A:Molecule type: mRNA
A:Residues: 6-79, '81-84, '86-381 <MED>
A:Cross-references: GB:M15799; NID:g181462; PIDN:AAA52167.1; PID:g181463
R:Moran, P.; Raab, H.; Kohn, W.J.; Caras, I.W.
J. Biol. Chem. 266, 1250-1257, 1991
A:Title: Glycophospholipid membrane anchor attachment. Molecular analysis of the cleavag
A:Reference number: A39101; MUID:91093238
A:Accession: A39101

A:Molecule type: protein
A:Residues: 338-352 <MOR>
R:Tublin, D.M.; Mallinson, G.; Poole, J.; Reid, M.E.; Thompson, E.S.; Ferdman, B.R.;
Blood 84, 1276-1282, 1994
A:Title: Molecular basis of reduced or absent expression of decay-accelerating factor
A:Reference number: I52594; MUID:94325573
A:Accession: I52594
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 194-198, 'L', 200-209 <LUB>
A:Cross-references: GB:S72858; NID:g639599; PIDN:AAC60633.1; PID:g639600
A:Experimental source: individual KW, Cromer blood group phenotype Dr(a-)
A:Note: the single nucleotide difference in this allele, which changes Ser-199 to Leu
on (see reference I52564), and thus reduced DAF expression
R:Reid, M.E.; Mallinson, G.; Sim, R.B.; Poole, J.; Pausch, V.; Merry, A.H.; Liew, Y.W
Blood 78, 3291-3297, 1991
A:Title: Biochemical studies on red blood cells from a patient with the Inab phenotype
A:Reference number: I52564; MUID:92075980
A:Accession: I52564
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 190-193, 'QLCPVE' <RE2>
A:Cross-references: GB:S70688; NID:g240301; PIDN:AAB20576.1; PID:g240302
A:Experimental source: individual KW, Cromer blood group phenotype Dr(a-) (described
C:Comment: Cromer blood group system antigens reside on this protein.
C:Comment: for an alternative splice form, see PIR:A26359.
C:Genetics:
A:Gene: GDB:DAF
A:Cross-references: GDB:I19088; OMIM:125240
A:Map position: lq32-lq32
C:Function:
A:Description: protects tissues from damage by regulating complement activation on ce
C:Superfamily: decay-accelerating factor; complement factor H repeat homology
C:Keywords: alternative splicing; blocked carboxyl end; complement inhibitor; glycopr
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-353/Product: decay-accelerating factor 2 #status predicted <MAT>
F:36-94/Domain: complement factor H repeat homology <FH01>
F:98-158/Domain: complement factor H repeat homology <FH02>
F:163-220/Domain: complement factor H repeat homology <FH03>
F:225-283/Domain: complement factor H repeat homology <FH04>
F:354-381/Domain: carboxyl-terminal propeptide #status predicted <PRO>
F:95/Binding site: carboxylate (Asn) (covalent) #status predicted
F:353/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature f

Query Match 22.4%; Score 261; DB 1; Length 381;
Best Local Similarity 29.9%; Pred. No. 1.5e-14;
Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

QY 1 EDCN----ELPPRRNTEILTGSWSDDQTY-PEGTQAIYKCRPGYR---SLGNVIMVCRKEW 53
DB 92 EFCNRSCVEPTRLNSASLQPYITQNYFPVGTWVEYECRPGYRREPSLSPLKTLQNLKW 151
QY 54 VALNPLRKCQRPCHGPDTPGTTFLTGGNVFEGVAVYTCNGYQLLGEINRECDT 113
DB 152 --STAVEFCKKSCPNPGEIRNGQIDVPGGILF--GATISFCSCNTGYKLFGSTS-SFCLI 206
QY 114 DG----WTNDIPICEVVKCLPVTAPENKIVSSAMEPDRE-YHFGOAVRFVCSNGYKIEG 168
DB 207 SSSVQWSDPLDECRYYC-----PAPPQIDNGIIGENDHYGROSVTYATKNGKFTMIG 261
QY 169 DEEMHC---SDGFWMSKEKPKC-----VETISCKSPDVIN-----GSPISQK 206
DB 262 EHSIYCTVNNDGSENGSGPPPECRGKSLTSKVPTTVQKPTTVNVPTEVSPTSQK 315

RESULT 9
A26359
decay-accelerating factor, splice form 1 precursor - human
N:Alternate names: decay-accelerating factor 5; decay-accelerating factor secreted fo
C:Species: Homo sapiens (man)
C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 02-Jun-2000
C:Accession: A26359; A39702; SI6187; S23138; A27258

R:Caras, I.W.; Davitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.
Nature 325, 545-549, 1987
A:Title: Cloning of decay-accelerating factor suggests novel use of splicing to generate
A:Reference number: A26359; MUID:87115845
A:Accession: A26359
A:Molecule type: mRNA
A:Residues: 1-440 <CAR>
A:Cross-references: GB:M30142
R:Ewulonu, U.K.; Ravi, L.; Medof, M.E.
Proc. Natl. Acad. Sci. U.S.A. 88, 4675-4679, 1991
A:Title: Characterization of the decay-accelerating factor gene promoter region.
A:Reference number: A39702; MUID:91271256
A:Accession: A39702
A:Molecule type: DNA
A:Residues: 1-79, 'T', 81-104 <EWU>
A:Cross-references: GB:M64356
A:Note: the authors translated the codon AGT for residue 85 as Met
R:Nakano, Y.; Sugita, Y.; Ishikawa, Y.; Choi, N.H.; Tobe, T.; Tomita, M.
Biochim. Biophys. Acta 1074, 326-330, 1991
A:Title: Isolation of two forms of decay-accelerating factor (DAF) from human urine.
A:Reference number: S16187; MUID:91291869
A:Accession: S16187
A>Status: preliminary
A:Molecule type: protein
A:Residues: 35-47 <BIO>
R:Nakano, Y.; Sumida, K.; Kikuta, N.; Miura, N.H.; Tobe, T.; Tomita, M.
Biochim. Biophys. Acta 1116, 235-240, 1992
A:Title: Complete determination of disulfide bonds localized within the short consensus
A:Reference number: S23138; MUID:92305034
A:Accession: S23138
A>Status: preliminary
A:Molecule type: protein
A:Residues: 35-41;63-68;79-81;93-103;128-134;143-145;155-159;162-168;188-192;203-204;211-
J. Biochem. 100, 143-150, 1986
A:Title: Improved method for the isolation and preliminary characterization of human DAF
A:Reference number: A27258; MUID:87008461
A:Accession: A27258
A:Molecule type: protein
A:Residues: 35, 'X', 37, 'G', 39-51, 'P', 53-55, 'X', 57-58, 'X', 60-63 <SUG>
A:Note: Gly-35 and Leu-38 were also found
C:Comment: For an alternative splice form, see PIR:B26359
C:Genetics:
A:Gene: GDB:DAF
A:Cross-references: GDB:119088; OMIM:125240
A:Map position: lq32-lq32
C:Superfamily: decay-accelerating factor; complement factor H repeat homology
F:1-34/Domain: alternative splicing; glycoprotein
F:35-440/Product: decay-accelerating factor 1 #status predicted <SIG>
F:36-94/Domain: complement factor H repeat homology <HAT>
F:98-158/Domain: complement factor H repeat homology <FH01>
F:163-220/Domain: complement factor H repeat homology <FH02>
F:225-283/Domain: complement factor H repeat homology <FH03>
F:95/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 22.4%; Score 261; DB 2; Length 440;
Best Local Similarity 29.9%; Pred. No. 1.8e-14;
Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

QY 1 EDNC---ELPPRRNTTEILTGSWSDQY-PEGTOAIYKCRPGYR--SLGNVIMVCRKGEW 53
DB 92 EFCNRCEVPTRLNSASLQPYITQYFVGVVVEYECRPGYRREPSPKLTCLQNLKW 151
QY 54 VALNPLRKCKRCPCGHPGDTPEFTTLTGGNVFEYGVKAVYTCNEGYQLLGEINRECDT 113
DB 152 --STAVEFCKKSCPNPGEIRNGQIDVPGILF--GATISFSCNTGYKLFGSTS-SFCL 206
QY 114 DG-----WTNDIPICEVVKCLPTAPENKIVSSAMEPDRE-YHFGQAVRFVNSGYKIEG 168
DB 207 SGGSVQWSDPLPECREIYC-----PAPPQIDNGIIQGERDHYGROSVTYACNKGFTMIG 261

QY 169 DEEMHC---SDDGFWMSKEKPKC-----VEISCKSPDVIN-----GSPISQK 206
DB 262 EHSIYCTVNNDEGWSGPPPECRCKSLTSKVPTVQKPTTVNPTVTEVSPTSQK 315
RESULT 10
A45900
complement C3b receptor type 2 long form precursor - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C:Accession: A45900; I48306
R:Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.
J. Immunol. 144, 3581-3591, 1990
A:Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2
A:Reference number: A45900; MUID:90229754
A:Accession: A45900
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-676 <KUR>
A:Cross-references: GB:M36470
A:Experimental source: clone 31-1
R:Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers,
J. Exp. Med. 181, 151-159, 1995
A:Title: Mouse complement regulatory protein Crry/p65 uses the specific mechanisms of
A:Reference number: I48306; MUID:95105691
A:Accession: I48306
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 21-367 <RES>
A:Cross-references: EMBL:U17128; NID:g595980; PIDN:AAA78271.1; PID:g595982
C:Genetics:
A:Gene: Cr2
A:Introns: 80/1; 113/2; 142/1; 275/1; 333/1
C:Superfamily: complement factor H repeat homology
F:22-78/Domain: complement factor H repeat homology <FH01>
F:83-140/Domain: complement factor H repeat homology <FH02>
F:145-211/Domain: complement factor H repeat homology <FH03>
F:217-272/Domain: complement factor H repeat homology <FH04>
F:276-331/Domain: complement factor H repeat homology <FH05>
F:336-394/Domain: complement factor H repeat homology <FH06>
F:399-458/Domain: complement factor H repeat homology <FH07>
F:467-523/Domain: complement factor H repeat homology <FH08>
F:531-587/Domain: complement factor H repeat homology <FH09>
F:592-648/Domain: complement factor H repeat homology <FH10>

Query Match 22.4%; Score 260; DB 2; Length 676;
Best Local Similarity 28.4%; Pred. No. 3.4e-14;
Matches 56; Conservative 45; Mismatches 72; Indels 24; Gaps 9;

QY 20 SDQT-YPEGTQAIYKCRPGYRSLGNVIMVCRKGEWALNPLRKCKRCPCGHPGDTPEGTF 78
DB 37 SDRSEFAIGTTWEYKCRPGYFRKSFITCLETSKW--SDAQOFCRKPCNMQPRLHGSV 94
QY 79 TLTGGNVFEYGVKAVYTCNEGYQLLGEINRECDTDTGWTNDIPICEVVKCLPTAP 134
DB 95 HINTG--IEFGSTIYSCNQGYRLIGD--SSATCTVSDNTVMWDMNPLCESIPSPAI 151
QY 135 ENKIVSSAMEPDREYHFGQAVRFVNSG-----YKIEGDEEMHCSDD----GFWSKEK 184
DB 152 SNQDFTSSSKD----SFYGMVVTYCHTKGNREKFLDLVGEKSIYCTSKDNQGVINWSP 208
QY 185 PKCV-EISCKSPDVING 200
DB 209 PQCIPIRVKCPMPIENG 225
RESULT 11
I73012
complement C3b/C4b receptor, membrane-bound form precursor - human
N:Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CR1); sur
N:Contains: complement C3b/C4b receptor, secreted form
C:Species: Homo sapiens (man)

C;Date: 02-Jul-1996 #sequence_revision 24-Nov-1999 #text_change 21-Jul-2000
C;Accession: I73012; I56203; A47602; S03291; S03843; A28507; A24748; B24748; C24748
R;Vik, D.P.; Wong, W.W. 1993
J. Immunol. 151, 6214-6224, 1993
A;Title: Structure of the gene for the F allele of complement receptor type 1 and sequen
A;Reference number: I56203; MUID:94065175
A;Accession: I73012
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-683, 'X', 685-894, 'A', 896-1000, 1451-1471, 'X', 1473-2489 <VIK1>
A;Cross-references: GB:L17418; NID:g306678; PIDN:AAB60695.1; PID:g451303
A;Accession: I56203
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-683, 'X', 685-894, 'A', 896-1000, 1451-1471, 'X', 1473-2489 <VIK2>
A;Cross-references: GB:L17418; NID:g306678; PIDN:AAB60694.1; PID:g306680
R;Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.; Wi
J. Exp. Med. 169, 847-863, 1989
A;Title: Structure of the human CRI gene. Molecular basis of the structural and quantita
A;Reference number: A47602; MUID:89176869
A;Accession: A47602
A;Molecule type: DNA
A;Residues: 1-41 <WON>
A;Cross-references: GB:X14893
R;Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.
J. Exp. Med. 168, 1255-1270, 1988
A;Title: Identification of an alternative polyadenylation site in the human C3b/C4b rece
type 1.
A;Reference number: S03291; MUID:89010527
A;Accession: S03291
A;Molecule type: mRNA
A;Residues: 26-584 <HOU>
A;Cross-references: EMBL:X1362; NID:g30197; PIDN:CAA32541.1; PID:g736240
A;Experimental source: Clone CRI-4
R;Klickstein, L.B.; Bartow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.
J. Exp. Med. 168, 1699-1717, 1988
A;Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4b r
A;Reference number: S03843; MUID:89035992
A;Accession: S03843
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-894, 'A', 896-1000, 1451-2064, 'I', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T', 2327-2
A;Cross-references: EMBL:Y00816; NID:g30185; PIDN:CAA60755.1; PID:g30186
R;Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Wells, J.H.; Wilson, J.G.; Fearon, D.T.
J. Exp. Med. 165, 1095-1112, 1987
A;Title: Human C3b/C4b receptor (CRI). Demonstration of long homologous repeating domain
A;Reference number: A28507; MUID:87168191
A;Accession: A28507
A;Molecule type: mRNA
A;Residues: 953-1221, 'FV', 1224-2064, 'I', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T', 2327-2
A;Cross-references: GB:X05309; NID:g30196; PIDN:CAA28933.1; PID:g809019
R;Wong, W.W.; Klickstein, L.B.; Smith, J.A.; Wells, J.H.; Fearon, D.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985
A;Title: Identification of a partial cDNA clone for the human receptor for complement fr
A;Reference number: A94073; MUID:86067975
A;Accession: A24748
A;Molecule type: mRNA
A;Residues: 311-333; 729-745; 831-845 <W02>
A;Cross-references: GB:M11569; NID:g180991; PIDN:AAA52297.1; PID:g180995; GB:M11617; NID
C;Genetics:
A;Gene: GDB:CD35
A;Cross-references: GDB:I19800; OMIM:120620
A;Map position: lq32-lq32
A;Introns: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2; 6
1484/2; 1513/1; 1646/1; 1705/1; 1739/2; 1768/1; 1844/1; 1904/1; 1937/2; 1965/1; 2059/1
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C;Keywords: duplication; glycoprotein; transmembrane protein
F;1-41/Domain: signal sequence #status: predicted <SIG>
F;42-2489/Product: complement C3b/C4b receptor, membrane-bound form #status predicted <
F;42-584/Product: complement C3b/C4b receptor, secreted form #status predicted <MATS>
F;43-99/Domain: complement factor H repeat homology <PH01>
F;104-161/Domain: complement factor H repeat homology <PH02>
F;166-232/Domain: complement factor H repeat homology <PH03>

F:238-293/Domain: complement factor H repeat homology <PH04>
F:297-353/Domain: complement factor H repeat homology <PH05>
F:358-416/Domain: complement factor H repeat homology <PH06>
F:421-487/Domain: complement factor H repeat homology <PH07>
F:493-549/Domain: complement factor H repeat homology <PH08>
F:554-611/Domain: complement factor H repeat homology <PH09>
F:616-682/Domain: complement factor H repeat homology <PH10>
F:688-743/Domain: complement factor H repeat homology <PH11>
F:747-803/Domain: complement factor H repeat homology <PH12>
F:808-866/Domain: complement factor H repeat homology <PH13>
F:871-937/Domain: complement factor H repeat homology <PH14>
F:943-999/Domain: complement factor H repeat homology <PH15>
F:1004-1061/Domain: complement factor H repeat homology <PH16>
F:1066-1132/Domain: complement factor H repeat homology <PH17>
F:1138-1193/Domain: complement factor H repeat homology <PH18>
F:1197-1253/Domain: complement factor H repeat homology <PH19>
F:1258-1317/Domain: complement factor H repeat homology <PH20>
F:1321-1387/Domain: complement factor H repeat homology <PH21>
F:1393-1449/Domain: complement factor H repeat homology <PH22>
F:1454-1511/Domain: complement factor H repeat homology <PH23>
F:1516-1582/Domain: complement factor H repeat homology <PH24>
F:1588-1643/Domain: complement factor H repeat homology <PH25>
F:1647-1703/Domain: complement factor H repeat homology <PH26>
F:1708-1766/Domain: complement factor H repeat homology <PH27>
F:1771-1837/Domain: complement factor H repeat homology <PH28>
F:1846-1902/Domain: complement factor H repeat homology <PH29>
F:1907-1964/Domain: complement factor H repeat homology <PH30>
F:1969-2035/Domain: complement factor H repeat homology <PH31>
F:2041-2096/Domain: complement factor H repeat homology <PH32>
F:2100-2156/Domain: complement factor H repeat homology <PH33>
F:2161-2219/Domain: complement factor H repeat homology <PH34>
F:2224-2290/Domain: complement factor H repeat homology <PH35>
F:2298-2354/Domain: complement factor H repeat homology <PH36>
F:2359-2415/Domain: complement factor H repeat homology <PH37>

Query Match 21.9%; Score 255; DB 2; Length 2489;

Best Local Similarity 30.3%; Pred. No. 3.6e-13;

Matches 67; Conservative 39; Mismatches 83; Indels 32; Gaps 13;

QY 3 CNE---LPPRRNTEILTCGWSQDTPEGTQAIYKCRPGYRSLGNVIMVCRKEWVALNPL 59

DB 43 CNAPELWLPFARNTNLT---DEFEPPGTYLYNECRPGYSGRPFSTICUKNSVMTGAKD- 97

QY 60 RKQKRCPCGHPGDTFTFTLTGGNVFVGKAVTCTGEGYQLLGEINRYREC----DTGG 115

DB 98 -RCRRKSCRNPDPVNGMVHVIKG--IQFGSQIKVYSGTKGYRLIGS-SSATCLISGDTVI 153

QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYHFGQAVRVCNSG-----YKIEG 168

DB 154 WDNETPICDRIPCGLPPTI-TNGDFISTNRE---NFHYGSVVYTRCNPGSGRKKVPELVG 209

QY 169 DEEMHC--SDD--GFMSKEKPKC-VEISCKSPDVINGSPIS 204

DB 210 EPSIYCTSNDDQVGIMSGPAPQCIIPNCKTPPNVENGILVS 250

RESULT 12

JC5092

E-selectin - pig

C;Species: Sus scrofa domestica (domestic pig)

C;Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 19-May-2000

C;Accession: JC5092

R;Winkler, H.; Brostjan, C.; Csizmadia, V.; Natarajan, G.; Anrather, J.; Bach, F.H.

Gene 176, 67-72, 1996

A;Title: The intron-exon structure of the porcine E-selectin-encoding gene.

A;Reference number: JC5092; MUID:97075911

A;Contents: endothelial cells

A;Accession: JC5092

A;Molecule type: DNA

A;Residues: 1-482 <WIN>

A;Cross-references: GB:U37521; NID:g1052974; PIDN:AAC48680.1; PID:g1052975

C;Comment: This protein is a member of the selectin family of adhesion molecules.

C:Genetics:

A: Introns: 13/1; 1421; 178/1; 237/1; 300/1; 363/1; 422/1; 459/1; 466/1
C: Superfamily: unassigned EGF-related proteins; C-type lectin homology; complement factor
F: 13-139/Domain: C-type lectin homology <LCH>
F: 181-235/Domain: complement factor H repeat homology <FH1>
F: 240-298/Domain: complement factor H repeat homology <FH2>
F: 303-361/Domain: complement factor H repeat homology <FH3>
F: 366-420/Domain: complement factor H repeat homology <FH4>

Query Match 21.8%; Score 253.5; DB 2; Length 482;
Best Local Similarity 28.6%; Pred. No. 8.3e-14;
Matches 56; Conservative 36; Mismatches 83; Indels 21; Gaps 9;

QY 22 QTYPEQTAIYKCRPGYRSLGNIMVC-RKGEWVALNPLRKCOKRPG---CHPGDTPFGT 77

Db 196 QSLPWNTTCAFECKEGELIGPEHLQCTSSGWDGKKP--TCRAVTCDTVGHQPON---GD 250

QY 78 FTLTGGNVFVGVKAV--YTCNKGYYQLLGEINRECDTDG-WTNDIPICEVWKLPLVAP 134

Db 251 VSNHSSIGEFAYKSTCHFTCAEGFGLGQPAQI-ECTAQGOWTQQAQVCKAVKCPAVSQP 309

QY 135 ENKGIYSSAMEPDREYHFGQAVRVCNSGYKIEGDEEMHCSDDGFWSKPKVCVEISCKS 194

Db 310 KNG-LVKFTHSPTEFTYKSCAFSCGEGFELGRSAQLACTSOGWTQVPSQVOCSS 368

QY 195 PDV-----INGSPI 203

Db 369 LEVPREINMSCGEPV 384

RESULT 13

I56234

decay-accelerating factor - orangutan (fragment)

C: Species: Pongo pygmaeus (orangutan)

C: Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Aug-1999

C: Accession: I56234

R: Nickells, M.W.; Alvarez, J.I.; Lublin, D.M.; Atkinson, J.P.

J. Immunol. 152, 676-685, 1994

A: Title: Characterization of DAF-2, a high molecular weight form of decay-accelerating

A: Reference number: I56234; MUID: 94110622

A: Accession: I56234

A: Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: mRNA

A: Residues: 1-340 <RES>

A: Cross-references: GB:S67775; NID:q459389; PID: AAC0609.1; PID: q459390

C: Superfamily: decay-accelerating factor; complement factor H repeat homology

F: 1-53/Domain: complement factor H repeat homology (fragment) <FH01>

F: 57-117/Domain: complement factor H repeat homology <FH02>

F: 122-179/Domain: complement factor H repeat homology <FH03>

F: 184-242/Domain: complement factor H repeat homology <FH04>

Query Match 21.6%; Score 251; DB 2; Length 340;
Best Local Similarity 29.6%; Pred. No. 9.3e-14;
Matches 69; Conservative 37; Mismatches 91; Indels 36; Gaps 12;

QY 1 EDGN---ELPPRNTTILGWSDDQY-PEGTAIYKCRPGYR---SLGNVIMVCKRGEW 53

Db 51 EFCNRSEVPTRNLASLAKQPYITQNYFPVGTVEYVCRPGRRELSLSKTLCLQLTW 110

QY 54 VALNPLRKCKRCPGHPGDTPTFTLTGNNVFEYGVKAVYTCNKGYYQLLGEINRECDT 113

Db 111 --STAVEFCKKSCNPGEIRNGIQDVSNGILF--GATISFCNTQYKLGFGPTS-SLCLI 165

QY 114 DG----WTNDIPICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRVCNSGYKIEG 169

Db 166 SGSSVOWSDPLPCREIYCPAPPOIDNG-IIQKRD---HYGYRQSIYACNKGTYWIGE 221

QY 170 EEMHCS---DDGFWSEKPKC-----VEISCKSPDVIN-----GSPISQK 206

Db 222 HSIYCTVNDDEGWSGPPPCRGKSLTSKVPTVQKPTTVNPTTEVSPTSQK 274

RESULT 14

I36936

complement receptor 1 - chimpanzee (fragment)

C: Species: Pan troglodytes (chimpanzee)

C: Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000

C: Accession: I36936; I36937

R: Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.

J. Immunol. 153, 691-700, 1994

A: Title: Primary sequence of an alternatively spliced form of CRI. Candidate for the

A: Reference number: I36935; MUID: 94292799

A: Accession: I36936

A: Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: mRNA

A: Residues: 1-2014 <BIR1>

A: Cross-references: GB:L24920; NID: g551564; PID: AAA51438.1; PID: g557725

A: Accession: I36937

A: Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: mRNA

A: Residues: 1-397,1751-2014 <BIR2>

A: Cross-references: GB:L24921; NID: g557726; PID: AAA51439.1; PID: g557727

C: Gene: CRI

C: Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology

C: Keywords: duplication; glycoprotein; tandem repeat

F: 18-74/Domain: complement factor H repeat homology <FH01>

F: 79-136/Domain: complement factor H repeat homology <FH02>

F: 141-207/Domain: complement factor H repeat homology <FH03>

F: 213-268/Domain: complement factor H repeat homology <FH04>

F: 272-328/Domain: complement factor H repeat homology <FH05>

F: 333-391/Domain: complement factor H repeat homology <FH06>

F: 396-462/Domain: complement factor H repeat homology <FH07>

F: 470-526/Domain: complement factor H repeat homology <FH08>

F: 529-586/Domain: complement factor H repeat homology <FH09>

F: 722-778/Domain: complement factor H repeat homology <FH10>

F: 1041-1107/Domain: complement factor H repeat homology <FH11>

F: 1172-1228/Domain: complement factor H repeat homology <FH12>

F: 1233-1291/Domain: complement factor H repeat homology <FH13>

F: 1296-1362/Domain: complement factor H repeat homology <FH14>

F: 1432-1489/Domain: complement factor H repeat homology <FH15>

F: 1625-1681/Domain: complement factor H repeat homology <FH16>

F: 1749-1815/Domain: complement factor H repeat homology <FH17>

F: 1823-1879/Domain: complement factor H repeat homology <FH18>

F: 1884-1940/Domain: complement factor H repeat homology <FH19>

Query Match 21.2%; Score 247; DB 2; Length 2014;
Best Local Similarity 29.9%; Pred. No. 1.3e-12;
Matches 66; Conservative 39; Mismatches 84; Indels 32; Gaps 13;

QY 3 CNE---LPPIRRTTEILTGWSDDQY-PEGTAIYKCRPGYRSLGNVIMVCKRGEWVALNPL 59

Db 18 CNAPEWLPFARPPTNLT---DEFEFGITGLNYECRPGYGRPFSLCLNSVWTGAKD- 72

QY 60 RKQKRCPCGHPGDTPTFTLTGNNVFEYGVKAVYTCNKGYYQLLGEINREYREC-----DTDG 115

Db 73 -RCRRSCRRNPDPVNGVHVI--KDIQFGSKYKISCTGYRLIGS-SSATCIISGDTVI 128

QY 116 WTNDIPICEVVKC-LPVTAPENKIVSSAMEPDREYHFGQAVRVCNSG-----YKIEG 168

Db 129 WDNETPICDIRICGLPPTI--TNGDFISTNRE---NFHYGSVVTYRCNPGSGGRKRVFELVG 184

QY 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDVINGSPIS 204

Db 185 EPSIYCTSDNDQVQGWNGPAPQCIIPNKCTPPNVENGILYS 225

RESULT 15

MWVZSP

apolipoprotein H homolog precursor - vaccinia virus

N: Alternate names: 35K secretory protein; C3L protein; virokin

C: Species: vaccinia virus

A: Note: host Homo sapiens (man)

C:Date: 31-Dec-1989 #sequence_revision 30-Jun-1990 #text_change 22-Jun-1999
C:Accession: A31005; B42504
R:Kotwal, G.J.; Moss, B.
Nature 335, 176-178, 1988
A:Title: Vaccinia virus encodes a secretory polypeptide structurally related to complement C3
A:Reference number: A31005; MUID:88318974
A:Accession: A31005
A:Molecule type: DNA
A:Residues: 1-263 <KOT>
A:CROSS-references: GB:X13166; NID:g60690; PIDN:CAA31564.1; PID:g60691
A:Experimental source: strain WR
R:Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 517-563, 1990
A:Title: Appendix to "The complete DNA sequence of vaccinia virus".
A:Reference number: A42501
A:Accession: B42504
A:Molecule type: DNA
A:Residues: 1-263 <GOE>
A:CROSS-references: GB:M35027; NID:g335317; PIDN:AAA47997.1; PID:g335345
A:Experimental source: strain Copenhagen
R:Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 247-266, 1990
A:Title: The complete DNA sequence of vaccinia virus.
A:Reference number: A42531; MUID:91021027
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolog
C:Keywords: duplication; extracellular protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-263/Product: C4b-binding protein homolog #status predicted <MAT>
F:21-81/Domain: complement factor H repeat homology <FH1>
F:86-143/Domain: complement factor H repeat homology <FH2>
F:148-201/Domain: complement factor H repeat homology <FH3>
F:206-261/Domain: complement factor H repeat homology <FH4>

Query Match 21.2%; Score 246; DB 1; Length 263;
Best Local Similarity 33.7%; Pred. No. 1.9e-13;
Matches 63; Conservative 23; Mismatches 83; Indels 18; Gaps 9;
QY 20 SDQTYPEGTQAIYKRCGYR--SLGNVIMVCRKGEWALNLRKQCRKRGHPGDPFGT 77
Db : : | | | | | : : | | | | | : : | | | | | : : | | | | |
40 ANANYNIGDTIEYLCPLGPKYKQKMGPIYAKCTGTGTLFN---QCIRKRCPSPRDIDNGQ 96
QY 78 FTLTGGNVFEYGVRAVYTCNEGYOLLGEI-NVRECDDTG---WTNDIPICEVVKCLPVTA 133
Db : : | | | | | : : | | | | | : : | | | | | : : | | | | |
97 LDIGG---VDFGSSITYSCNSGYHLIGESKSYCELGSTGSMWNPEADICESVKQSPPS 153
QY 134 PENGKIVSSAMEPDREYHFGQAVRFVCMNSGYKIEGDEEMHCGDDGFWSKEPKVCVEISCK 193
Db : : | | | | | : : | | | | | : : | | | | | : : | | | | |
154 ISNGR--HNGYED--FYTDGSVVTVSCNSGYSLIGNSGVLCS-GGEWS-DPPTCOIVKCP 207
QY 194 SPDVING 200
Db : : | |
208 HPTISNG 214

Search completed: August 29, 2002, 15:02:15
Job time: 86 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:02:24 ; Search time 21.93 Seconds
(without alignments)
365.479 Million cell updates/sec

Title: US-09-316-163-9

Perfect score: 1163
Sequence: 1 EDNCLEPPRRTEILTGWS.....VEISCKSPDVINGSPISQKI 207

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1163	100.0	1231	1 CFAH_HUMAN	P08603 homo sapien
2	822	70.7	1234	1 CFAH_MOUSE	P06909 mus musculus
3	523.5	45.0	685	1 CFAH_BOVIN	Q28085 bos taurus
4	261	22.4	381	1 DAF_HUMAN	P08174 homo sapien
5	255	21.9	2039	1 CRL_HUMAN	P17927 homo sapien
6	253.5	21.8	484	1 LEM2_FIG	P98110 sus scrofa
7	251	21.6	340	1 DAF_PONPY	P49457 pongo pygma
8	246.5	21.2	507	1 DAF_CAVPO	Q60401 cavia porce
9	246	21.2	263	1 VCP_VACCV	P10998 vaccinia vi
10	244	21.0	597	1 C4BP_HUMAN	P04003 homo sapien
11	238.5	20.5	377	1 MCP_HUMAN	P15529 homo sapien
12	236	20.3	360	1 CDPH_HSVSA	Q01016 herpesvirus
13	231.5	19.9	390	1 DAF1_MOUSE	Q61475 mus musculus
14	230.5	19.8	610	1 LEM2_HUMAN	P16581 homo sapien
15	228	19.6	469	1 C4BP_MOUSE	P08607 mus musculus
16	227.5	19.6	768	1 LEM3_MOUSE	Q01102 mus musculus
17	227.5	19.6	830	1 LEM3_HUMAN	P16109 homo sapien
18	226	19.4	549	1 LEM2_RAT	P98105 rat mus norv
19	222.5	19.1	407	1 DAF2_MOUSE	Q61476 mus musculus
20	222	19.1	768	1 LEM3_RAT	P98106 rat mus norv
21	220	18.9	611	1 LEM2_CANFA	P33730 canis fami
22	218.5	18.8	612	1 LEM2_MOUSE	Q00690 mus musculus
23	216	18.6	558	1 C4BP_RAT	Q63514 rat mus norv
24	214.5	18.4	551	1 LEM2_RABIT	P27113 oryctolagus
25	214.5	18.4	610	1 C4BP_BOVIN	Q28065 bos taurus
26	214	18.4	769	1 LEM3_SHEEP	P98109 ovis aries
27	212	18.2	1025	1 CR2_MOUSE	P19070 mus musculus
28	210.5	18.1	646	1 LEM3_BOVIN	P42201 bos taurus
29	210.5	18.1	1033	1 CR2_HUMAN	P20023 homo sapien
30	210	18.1	372	1 LEM1_MOUSE	P18337 mus musculus
31	205.5	17.7	485	1 LEM2_BOVIN	P98107 bos taurus
32	205	17.6	345	1 APOH_BOVIN	P17690 bos taurus
33	202	17.4	958	1 HITG_DROME	Q09101 drosophila

ALIGNMENTS

```

RESULT 1
CFAH_HUMAN
ID CFAH_HUMAN STANDARD; PRT: 1231 AA.
AC P08603; Q14570; P78435; Q9NU86;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement factor H precursor (H factor 1).
GN HFI OR HF OR CFH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT Y-402.
RC TISSUE=Liver;
RX MEDLINE=88134059; PubMed=2963625;
RA Ripoché J., Day A.J., Harris T.J.R., Sim R.B.;
RT "The complete amino acid sequence of human complement factor H.";
RL Biochem. J. 249:593-602(1988).
RN [2]
RP SEQUENCE OF 53-445 FROM N.A.
RX MEDLINE=87054207; PubMed=2946589;
RA Schulz T.F., Schwaible W., Stanley K.K., Dierich M.P.;
RT "Human complement factor H: isolation of cDNA clones and partial cDNA
sequence of the 38-kDa tryptic fragment containing the binding site
for C3b.";
RL Eur. J. Immunol. 16:1351-1355(1986).
RN [3]
RP SEQUENCE OF 226-449 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=86169701; PubMed=2937845;
RA Kristensen T., Wetsel R.A., Tack B.F.;
RT "Structural analysis of human complement protein H: homology with C4b
binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";
RL J. Immunol. 136:3407-3411(1986).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RL Bird C.;
RP Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1047-1231 FROM N.A.
RX MEDLINE=91201892; PubMed=1826708;
RA Estaller C., Koistinen V., Schwaible W., Dierich M.P., Weiss E.H.;
RT "Cloning of the 1.4-kb mRNA species of human complement factor H
reveals a novel member of the short consensus repeat family related
to the carboxy terminal of the classical 150-kDa molecule.";
RL J. Immunol. 146:3190-3196(1991).
RN [6]
RP SEQUENCE OF 19-35.
RX MEDLINE=83048213; PubMed=6215918;
RA Sim R.B., Discipio R.G.;
RT "Purification and structural studies on the complement-system control
protein beta 1H (Factor H)";
RL Biochem. J. 205:285-293(1982).
RN [7]
RP SEQUENCE OF 1-19 FROM N.A.

```

```

34 200 17.2 372 1 LEM1_RAT
35 199 17.1 345 1 APOH_MOUSE
36 194 16.7 372 1 LEM1_HUMAN
37 194 16.7 372 1 LEM1_MACMU
38 194 16.7 372 1 LEM1_PANTR
39 194 16.7 372 1 LEM1_PAPHA
40 193 16.6 372 1 LEM1_PONPY
41 191.5 16.5 661 1 F13B_HUMAN
42 188 16.2 258 1 C4BB_RAT
43 188 16.2 1019 1 LFC_TACTR
44 187.5 16.1 345 1 APOH_HUMAN
45 184 15.8 1019 1 LFC_CARRO
P30836 rattus norv
Q01339 mus musculus
P14151 homo sapien
Q95198 macaca mula
Q95237 pan troglod
Q28768 papio hamad
Q95235 pongo pygma
P05160 homo sapien
Q63515 rattus norv
P28175 tachypieus
P02749 homo sapien
Q26422 carcinoscor

```

RA	RL	RP	RR	RS	RT	RU	RV	RW	RX	RY	RZ	RA	RL	RP	RR	RS	RT	RU	RV	RW	RX	RY	RZ	RA	RL	RP	RR	RS	RT	RU	RV	RW	RX	RY	RZ	RA	RL	RP	RR	RS	RT	RU	RV	RW	RX	RY	RZ	RA	RL	RP	RR	RS	RT	RU	RV	RW	RX	RY	RZ																																																																																																																																																																																																																																																																																																																
Vik D.P., Williams S.A.;	Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.	[8]	SEQUENCE OF 1-9 FROM N.A.	Dominguez O.;	Thesis (1993), Hospital Trias I Pujol, Spain.	[9]	STRUCTURE BY NMR OF 927-985 (SUSHI 16).	MEDLINE-91278097; PubMed-1829116;	Norman D.G., Barlow P.N., Sim B., Campbell I.D.;	"Three-dimensional structure of a complement control protein module in solution.";	J. Mol. Biol. 219:717-725(1991).	[10]	STRUCTURE BY NMR OF 264-322 (SUSHI 5).	MEDLINE-9232649; PubMed-1533152;	Barlow P.N., Norman D.G., Steinkasserer A., Horne T.J., Pearce J., Driscoll P.C., Sim B., Campbell I.D.;	"Solution structure of the fifth repeat of factor H: a second example of the complement control protein module.";	Biochemistry 31:3626-3634(1992).	[11]	STRUCTURE BY NMR OF 866-985 (SUSHI 15 AND 16).	MEDLINE-93333119; PubMed-8331663;	Barlow P.N., Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P., Sim B., Campbell I.D.;	"Solution structure of a pair of complement modules by nuclear magnetic resonance.";	J. Mol. Biol. 232:268-284(1993).	CC	-1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5 CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.	CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; ARE PRODUCED BY ALTERNATIVE SPLICING.	CC	-1- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.	CC	-1- CAUTION: REF 2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 341 ONWARD DUE TO A FRAMESHIFT.	CC	-----	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	CC	-----	CC	EMBL; Y00716; CAA68704.1; -	DR	EMBL; X04697; CAB41739.1; ALT_FRAME.	DR	EMBL; X07523; CAA30403.1; -	DR	EMBL; M12383; AAA52013.1; -	DR	EMBL; AL049744; CAB70598.1; -	DR	EMBL; M65294; AAA35948.1; -	DR	EMBL; U56979; AAB01987.1; -	DR	EMBL; Z29665; CAA82763.1; -	DR	PIR; S00254; NBHUH.	DR	PIR; S03013; S03013.	DR	PDB; 1HCC; 15-APR-92.	DR	PDB; 1HFH; 15-JUL-93.	DR	PDB; 1HFI; 15-JUL-93.	DR	MIM; 134370; -	DR	InterPro: IPR000436; Sush1_SCR_CCP.	DR	Pfam; PF000084; sush1; 20.	DR	SMART; SM00032; CCP; 20.	DR	Complete alternate pathway; Plasma; Glycoprotein; Repeat; Sush1; Signal; 3D-structure; Polymorphism; Alternative splicing.	KW	SIGNAL	1	18	FT	CHAIN	19	1231	COMPLEMENT FACTOR H.	FT	DOMAIN	20	81	SUSHI 1.	FT	DOMAIN	84	142	SUSHI 2.	FT	DOMAIN	145	206	SUSHI 3.	FT	DOMAIN	209	263	SUSHI 4.	FT	DOMAIN	266	321	SUSHI 5.	FT	DOMAIN	321	376	SUSHI 6.	FT	DOMAIN	376	431	SUSHI 7.	FT	DOMAIN	431	486	SUSHI 8.	FT	DOMAIN	486	541	SUSHI 9.	FT	DOMAIN	541	596	SUSHI 10.	FT	DOMAIN	596	651	SUSHI 11.	FT	DOMAIN	651	706	SUSHI 12.	FT	DOMAIN	706	761	SUSHI 13.	FT	DOMAIN	761	816	SUSHI 14.	FT	DOMAIN	816	871	SUSHI 15.	FT	DOMAIN	871	926	SUSHI 16.	FT	DOMAIN	926	981	SUSHI 17.	FT	DOMAIN	981	1036	SUSHI 18.	FT	DOMAIN	1036	1091	SUSHI 19.	FT	DOMAIN	1091	1146	SUSHI 20.	FT	DOMAIN	1146	1201	BY SIMILARITY.	FT	DISULFID	21	66	BY SIMILARITY.	FT	DISULFID	52	80	BY SIMILARITY.	FT	DISULFID	85	129	BY SIMILARITY.	FT	DISULFID	114	141	BY SIMILARITY.	FT	DISULFID	146	192	BY SIMILARITY.	FT	DISULFID	178	205	BY SIMILARITY.	FT	DISULFID	210	251	BY SIMILARITY.	FT	DISULFID	237	262	BY SIMILARITY.	FT	DISULFID	267	309	BY SIMILARITY.	FT	DISULFID	294	320	BY SIMILARITY.	FT	DISULFID	325	374	BY SIMILARITY.	FT	DISULFID	357	385	BY SIMILARITY.	FT	DISULFID	389	431	BY SIMILARITY.	FT	DISULFID	416	442	BY SIMILARITY.	FT	DISULFID	448	494	BY SIMILARITY.	FT	DISULFID	477	505	BY SIMILARITY.	FT	DISULFID	509	553	BY SIMILARITY.	FT	DISULFID	536	564	BY SIMILARITY.	FT	DISULFID	569	611	BY SIMILARITY.	FT	DISULFID	597	623	BY SIMILARITY.	FT	DISULFID	630	673	BY SIMILARITY.	FT	DISULFID	659	684	BY SIMILARITY.	FT	DISULFID	691	733	BY SIMILARITY.	FT	DISULFID	719	744	BY SIMILARITY.	FT	DISULFID	753	792	BY SIMILARITY.	FT	DISULFID	781	803	BY SIMILARITY.	FT	DISULFID	811	853	BY SIMILARITY.	FT	DISULFID	839	864	BY SIMILARITY.	FT	DISULFID	870	915	BY SIMILARITY.	FT	DISULFID	901	926	BY SIMILARITY.	FT	DISULFID	931	973	BY SIMILARITY.	FT	DISULFID	959	984	BY SIMILARITY.	FT	DISULFID	989	1032	BY SIMILARITY.	FT	DISULFID	1018	1043	BY SIMILARITY.	FT	DISULFID	1048	1091	BY SIMILARITY.	

[illegible]

CC	EMBL; M12660; AAA37759.1; -	1	18	BY SIMILARITY.
DR	EMBL; J02891; AAA37795.1; -	19	1234	COMPLEMENT FACTOR H.
DR	EMBL; M31979; AAA37762.1; -	20	81	SUSHI 1.
DR	PIR; A26154; NBMSH.	84	142	SUSHI 2.
DR	HSSP; P08603; IHFI.	145	206	SUSHI 3.
DR	MGD; MGI:88385; Cfh.	209	263	SUSHI 4.
DR	InterPro; IPR000436; Sushi_SCR_CCP.	266	321	SUSHI 5.
DR	Pfam; PF00084; sushi; 20.	324	386	SUSHI 6.
DR	SMART; SM00032; CCP; 20.	388	443	SUSHI 7.
KW	Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;	447	506	SUSHI 8.
KW	Signal.	508	565	SUSHI 9.
FT	SIGNAL.	568	623	SUSHI 10.
FT	CHAIN	628	684	SUSHI 11.
FT	DOMAIN	689	744	SUSHI 12.
FT	DOMAIN	751	803	SUSHI 13.
FT	DOMAIN	807	862	SUSHI 14.
FT	DOMAIN	866	932	SUSHI 15.
FT	DOMAIN	935	990	SUSHI 16.
FT	DOMAIN	993	1049	SUSHI 17.
FT	DOMAIN	1052	1108	SUSHI 18.
FT	DOMAIN	1113	1169	SUSHI 19.
FT	DOMAIN	1171	1234	SUSHI 20.
FT	DISULFID	21	66	BY SIMILARITY.
FT	DISULFID	52	80	BY SIMILARITY.
FT	DISULFID	85	129	BY SIMILARITY.
FT	DISULFID	114	141	BY SIMILARITY.
FT	DISULFID	146	192	BY SIMILARITY.
FT	DISULFID	178	205	BY SIMILARITY.
FT	DISULFID	210	251	BY SIMILARITY.
FT	DISULFID	237	262	BY SIMILARITY.
FT	DISULFID	267	309	BY SIMILARITY.
FT	DISULFID	294	320	BY SIMILARITY.
FT	DISULFID	325	374	BY SIMILARITY.
FT	DISULFID	357	385	BY SIMILARITY.
FT	DISULFID	389	431	BY SIMILARITY.
FT	DISULFID	416	442	BY SIMILARITY.
FT	DISULFID	448	494	BY SIMILARITY.
FT	DISULFID	477	505	BY SIMILARITY.
FT	DISULFID	509	553	BY SIMILARITY.
FT	DISULFID	536	564	BY SIMILARITY.
FT	DISULFID	569	610	BY SIMILARITY.
FT	DISULFID	597	622	BY SIMILARITY.
FT	DISULFID	629	672	BY SIMILARITY.
FT	DISULFID	658	683	BY SIMILARITY.
FT	DISULFID	690	732	BY SIMILARITY.
FT	DISULFID	718	743	BY SIMILARITY.
FT	DISULFID	752	791	BY SIMILARITY.
FT	DISULFID	780	802	BY SIMILARITY.
FT	DISULFID	808	850	BY SIMILARITY.
FT	DISULFID	836	861	BY SIMILARITY.
FT	DISULFID	867	920	BY SIMILARITY.
FT	DISULFID	906	931	BY SIMILARITY.
FT	DISULFID	935	978	BY SIMILARITY.
FT	DISULFID	964	989	BY SIMILARITY.
FT	DISULFID	994	1037	BY SIMILARITY.
FT	DISULFID	1023	1048	BY SIMILARITY.
FT	DISULFID	1053	1096	BY SIMILARITY.
FT	DISULFID	1082	1107	BY SIMILARITY.
FT	DISULFID	1114	1157	BY SIMILARITY.
FT	DISULFID	1143	1168	BY SIMILARITY.
FT	DISULFID	1172	1223	BY SIMILARITY.

```
FT DISULFID 1206 1233 BY SIMILARITY.
FT CARBOHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 773 773 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1225 1225 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1234 AA; 139082 MW; CSAC02F341B957F7 CRC64;

Query Match 70.7%; Score 822; DB 1; Length 1234;
Best Local Similarity 68.0%; Pred. No. 3.7e-65;
Matches 140; Conservative 27; Mismatches 39; Indels 0; Gaps 0;

QY 1 EDNCNELPPRRNTTEILTGSDQTYPECTQAIYKCRPGYSLGNVIMVCRKGWVALNPLR 60
Db 19 EDCKGPPPRENSILSGNSWSEQLYPECTQATYKCRPGYRLTGTIVKCKNGKWWASPSR 78
QY 61 KCOKRPGCHPGDPFTTGTGNVFEYGVKAVYTCNEGYYQLLGEINRECDTGDWTNDI 120
Db 79 ICRKKPGCHPGDPFTFGSRLAVSQEFAGKVVYTCDDGYQLLGEIDYRECGADGWINDI 138
QY 121 PICEVVKCLPVTAPENCKIYSSAMEPDREYHFGQAVRFVCSNGYKKEGDEMHCSDGFW 180
Db 139 PLCEVVKCLPVTLENGRIVSGAETDQYVFGVVRFCNSGFKTEGKHEIHCSENGLW 198
QY 181 SKPKKVCISCKSPDVINGSPISQK 206
Db 199 SNEKPRCVELTCTPPRVENGGINVK 224

RESULT 3
ID CFAH_BOVIN STANDARD; PRT; 685 AA.
AC Q28085;
DC 01-MAR-2002 (rel. 41, Created)
DT 01-MAR-2002 (rel. 41, Last sequence update)
DE 01-MAR-2002 (rel. 41, Last annotation update)
DE Complement factor H (H factor 1) (Fragments).
GN HF1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 17-685 FROM N.A., AND SEQUENCE OF 1-16.
RC TISSUE=Liver;
RX MEDLINE=96202005; PubMed=8615824;
RA Soames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
RL in the interaction with complement component C3b."
RL Biochem. J. 315:523-531(1996).
CC -!- FUNCTION: Factor H functions as a cofactor in the inactivation of
CC C3b by factor I and also increases the rate of dissociation of the
CC C3bB complex (C3 convertase) and the (C3b)NBB complex (C5
CC convertase) in the alternative complement pathway (By similarity).
CC -!- SIMILARITY: CONTAINS AT LEAST 13 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X98697; CAA67257.1;
DR HSP; P10998; 1VVD.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sush1; 11.
DR SMART; SM00032; CCP; 11.
```

```
KW Complement alternate pathway; Plasma; Repeat; Sush1.
FT NON_TER 1 1
FT NON_CONS 16 17
FT DOMAIN <17 67 SUSHI 2.
FT DOMAIN 70 131 SUSHI 3.
FT DOMAIN 134 188 SUSHI 4.
FT DOMAIN 191 246 SUSHI 5.
FT DOMAIN 249 309 SUSHI 6.
FT DOMAIN 311 366 SUSHI 7.
FT DOMAIN 370 429 SUSHI 8.
FT DOMAIN 431 488 SUSHI 9.
FT DOMAIN 491 547 SUSHI 10.
FT DOMAIN 553 609 SUSHI 11.
FT DOMAIN 614 668 SUSHI 12.
FT DOMAIN 675 >685 SUSHI 13.
FT DISULFID 39 66 BY SIMILARITY.
FT DISULFID 71 117 BY SIMILARITY.
FT DISULFID 103 130 BY SIMILARITY.
FT DISULFID 135 176 BY SIMILARITY.
FT DISULFID 162 187 BY SIMILARITY.
FT DISULFID 192 234 BY SIMILARITY.
FT DISULFID 219 245 BY SIMILARITY.
FT DISULFID 250 297 BY SIMILARITY.
FT DISULFID 280 308 BY SIMILARITY.
FT DISULFID 312 354 BY SIMILARITY.
FT DISULFID 339 365 BY SIMILARITY.
FT DISULFID 371 417 BY SIMILARITY.
FT DISULFID 400 428 BY SIMILARITY.
FT DISULFID 432 476 BY SIMILARITY.
FT DISULFID 459 487 BY SIMILARITY.
FT DISULFID 492 534 BY SIMILARITY.
FT DISULFID 520 546 BY SIMILARITY.
FT DISULFID 554 597 BY SIMILARITY.
FT DISULFID 583 608 BY SIMILARITY.
FT DISULFID 615 656 BY SIMILARITY.
FT DISULFID 642 667 BY SIMILARITY.
FT NON_TER 685 685
SQ SEQUENCE 685 AA; 77536 MW; 69FC9DC8D530E872 CRC64;

Query Match 45.0%; Score 523.5; DB 1; Length 685;
Best Local Similarity 47.1%; Pred. No. 5.5e-39;
Matches 97; Conservative 21; Mismatches 31; Indels 57; Gaps 1;

QY 1 EDNCNELPPRRNTTEILTGSDQTYPECTQAIYKCRPGYSLGNVIMVCRKGWVALNPLR 60
Db 1 EDCKEPPPRKETILSVS----- 18
QY 61 KCOKRPGCHPGDPFTTGTGNVFEYGVKAVYTCNEGYYQLLGEINRECDTGDWTNDI 120
Db 19 -----GSPHLAEGNQFEYGAQVYVTCDEGYQYVGMNFRCDTNGWTNDI 63
QY 121 PICEVVKCLPVTAPENCKIYSSAMEPDREYHFGQAVRFVCSNGYKKEGDEMHCSDGFW 180
Db 64 PICEVVKCLPVTPEPKGFPSDALEPQDEYTVQGVVQFECNSGYMLDGPQIHCSSAGVW 123
QY 181 SKPKKVCISCKSPDVINGSPISQK 206
Db 124 SAEPTKCVEIFCKPPVILNQAVLPK 149

RESULT 4
ID DAF_HUMAN STANDARD; PRT; 381 AA.
AC P08174; P09679; P78361;
DT 01-AUG-1988 (rel. 08, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DT 01-MAR-2002 (rel. 41, Last annotation update)
DE Complement decay-accelerating factor precursor (CD55 antigen).
GN DAF OR CR OR CD55
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

NCBI_TaxID=9606;
 [1] SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RP MEDLINE=87115845; PubMed=2433596;
 RA Caras I.W., Davitz M.A., Rhee L., Weddell G., Martin D.W. Jr.,
 RA Nussenzweig V.;
 RT "Cloning of decay-accelerating factor suggests novel use of splicing
 RT to generate two proteins.";
 RL Nature 325:545-549(1987).
 [2] SEQUENCE FROM N.A.
 RP TISSUE=Cervix;
 RA Strausberg R.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 [3] SEQUENCE OF 6-381 FROM N.A. (ISOFORM 2).
 RX MEDLINE=87173602; PubMed=2436222;
 RA Medof M.E., Lublin D.M., Holers V.M., Ayers D.J., Getty R.R.,
 RA Leykam J.F., Atkinson J.P., Tykocinski M.L.;
 RT "Cloning and characterization of cDNAs encoding the complete sequence
 RT of decay-accelerating factor of human complement.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2007-2011(1987).
 [4] SEQUENCE OF 35-381 FROM N.A. (ISOFORM 2).
 RP TISSUE=Hippocampus;
 RC Kumar V.B., Hyung C., Nakta R., Walters M., Sasser T., Bernardo A.;
 RA "Decay-accelerating factor (DAF; CD 55) in the brain of Alzheimer's
 RT disease patients.";
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 [5] SEQUENCE OF 1-100 FROM N.A.
 RX MEDLINE=91271256; PubMed=1711208;
 RA Ewulonu U.K., Ravi L., Medof M.E.;
 RT "Characterization of the decay-accelerating factor gene promoter
 RT region.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4675-4679(1991).
 [6] SEQUENCE OF 35-46.
 RP TISSUE=Urine;
 RC MEDLINE=91291869; PubMed=1712233;
 RA Nakano Y., Sugita Y., Ishikawa Y., Choi N.-H., Tobe T., Tomita M.;
 RT "Isolation of two forms of decay-accelerating factor (DAF) from human
 RT urine.";
 RL Biochim. Biophys. Acta 1074:326-330(1991).
 [7] GPI-ANCHOR.
 RP MEDLINE=91093238; PubMed=1824699;
 RA Moran P., Raab H., Kohr W.J., Caras I.W.;
 RT "Glycophospholipid membrane anchor attachment. Molecular analysis of
 RT the cleavage/attachment site.";
 RL J. Biol. Chem. 266:1250-1257(1991).
 [8] DISULFIDE BONDS IN SUSHI DOMAINS.
 RP MEDLINE=92305034; PubMed=1377029;
 RA Nakano Y., Sumida K., Kikuta N., Miura N.-H., Tobe T., Tomita M.;
 RT "Complete determination of disulfide bonds localized within the short
 RT consensus repeat units of decay accelerating factor (CD55 antigen).";
 RL Biochim. Biophys. Acta 1116:235-240(1992).
 [9] FUNCTION AS A ECHOVIRUS RECEPTOR.
 RP MEDLINE=95045399; PubMed=7525274;
 RA Ward T., Pipkin P.A., Clarkson N.A., Stone D.M., Minor P.D.,
 RA Almond J.W.;
 RT "Decay-accelerating factor CD55 is identified as the receptor for
 RT echovirus 7 using CELFICS, a rapid immuno-focal cloning method.";
 RL EMBO J. 13:5070-5074(1994).
 [10] VARIANT BLOOD GROUP DR(A-).
 RP MEDLINE=94325573; PubMed=7519480;
 RA Lublin D.M., Mallinson G., Poole J., Reid M.E., Thompson E.S.,
 RA Ferman B.R., Telen M.J., Anstee D.J., Tanner M.J.A.;
 RT "Molecular basis of reduced or absent expression of
 RT decay-accelerating factor in Cromer blood group phenotypes.";

Blood 84:1276-1282(1994).
 -|- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
 CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
 CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
 CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES
 CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND
 CC FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS
 CC THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF
 CC THE COMPLEMENT CASCADE.
 CC -|- FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECHOVIRUS 7 AND RELATED
 CC VIRUSES (ECHOVIRUSES 13, 21, 29 AND 33).
 CC -|- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT
 CC HOMODIMER (MINOR FORM).
 CC -|- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -|- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/DAF-1 AND 2/DAF-2 (SHOWN
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -|- TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL
 CC TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA COMPONENT
 CC PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS
 CC Lining EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE
 CC ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.
 CC -|- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
 CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
 CC ACTIVE SITE ON SCR3 (BY SIMILARITY).
 CC -|- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.
 CC -|- POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CROMER BLOOD GROUP
 CC SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CR(A),
 CC TC(A), DR(A), ES(A), WES(B), UMC, AND IFC) AND LOW-INCIDENCE
 CC (TC(B), TC(C), AND WES(A)) ANTIGENS THAT RESIDE ON DAF. IN THE
 CC CROMER PHENOTYPES DR(A-) AND INAB THERE IS REDUCED OR ABSENT
 CC EXPRESSION OF DAF, RESPECTIVELY. IN THE CASE OF THE DR(A-)
 CC PHENOTYPE, A SINGLE NUCLEOTIDE SUBSTITUTION WITHIN EXON 5 ACCOUNTS
 CC FOR TWO CHANGES: A SIMPLE AMINO ACID SUBSTITUTION THAT IS THE
 CC BASIS OF THE ANTIGENIC VARIATION, AND AN ALTERNATIVE SPLICING
 CC EVENT THAT UNDERLIES THE DECREASED EXPRESSION OF DAF IN THIS
 CC PHENOTYPE.
 CC -|- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
 CC -|- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
 CC (RCA) FAMILY.
 CC -|- DATABASE: NAME=PROW; NOTE=CD guide CD55 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd55.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M31516; AAA52169.1; -;
 DR EMBL; M30142; AAA52168.1; -;
 DR EMBL; BC001288; AAH01288.1; -;
 DR EMBL; M15799; AAA52167.1; -;
 DR EMBL; U88576; ABA48622.1; -;
 DR EMBL; M64553; AAA52170.1; -;
 DR EMBL; M64356; AAA52170.1; JOINED.
 DR EMBL; S72858; AAC60633.1; -;
 DR PIR; B26359; B26359.
 DR PIR; A26359; A26359.
 DR PIR; S16187; S16187.
 DR PIR; A39101; A39101.
 DR PIR; S23138; S23138.
 DR HSSP; P08603; 1HCC.
 DR MIM; 125240; -;
 DR InterPro; IPR000436; Sush1_SCR_CCP.
 DR Pfam; PF00084; sush1; 4.
 DR SMART; SM00032; CCP; 4.
 KW Complement pathway; Plasma; Glycoprotein; Membrane; Repeat;
 KW Alternative splicing; GPI-anchor; Signal; Sushi; Polymorphism;
 KW Blood group antigen.
 FT SIGNAL 1 34
 FT CHAIN 35 353 COMPLEMENT DECAY-ACCELERATING FACTOR.


```
FT DOMAIN 357 417 SUSHI A6.
FT DOMAIN 420 488 SUSHI A7.
FT DOMAIN 492 550 SUSHI B1.
FT DOMAIN 553 612 SUSHI B2.
FT DOMAIN 615 683 SUSHI B3.
FT DOMAIN 687 744 SUSHI B4.
FT DOMAIN 746 804 SUSHI B5.
FT DOMAIN 807 867 SUSHI B6.
FT DOMAIN 870 938 SUSHI B7.
FT DOMAIN 942 1000 SUSHI C1.
FT DOMAIN 1003 1062 SUSHI C2.
FT DOMAIN 1065 1133 SUSHI C3.
FT DOMAIN 1137 1194 SUSHI C4.
FT DOMAIN 1196 1254 SUSHI C5.
FT DOMAIN 1257 1317 SUSHI C6.
FT DOMAIN 1320 1388 SUSHI C7.
FT DOMAIN 1395 1453 SUSHI D1.
FT DOMAIN 1456 1515 SUSHI D2.
FT DOMAIN 1518 1586 SUSHI D3.
FT DOMAIN 1590 1647 SUSHI D4.
FT DOMAIN 1649 1707 SUSHI D5.
FT DOMAIN 1710 1770 SUSHI D6.
FT DOMAIN 1773 1841 SUSHI D7.
FT DOMAIN 1847 1905 SUSHI E1.
FT DOMAIN 1908 1966 SUSHI E2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 338 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.
FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
```

```
FT DISULFID 1591 1633 BY SIMILARITY.
FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.
FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1739 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.
FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 959 959 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1028 1028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1310 1310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1481 1481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1540 1540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1605 1605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;
```

Query Match 21.9%; Score 255; DB 1: Length 2039;

Best Local Similarity 30.3%; Pred. No. 1,1,1,14; Matches 67; Conservative 39; Mismatches 83; Indels 32; Gaps 13;

```
OY 3 CNE---LPPRRNTEILTGSMSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGWVALNPL 59
Db 43 CNAPELWLPARTNLT---DEFFPPIGYLYNECRPGYSGRPFSLCLKNSVWTGAKD- 97
OY 60 RKCQRPCGHPGDTPEGTFTLTGGNVFVGVKAVYTCNKGYQLLGEINREC-----DTDG 115
Db 98 -RCRRKSCRNPDPVNGMVHVING--IOFGSQIKYSGCTGYRLIGS-SSATCIISGDTVI 153
OY 116 WTNDIPICEVVKC-LPVTAPENKIVSSAMEPDREYHFGQAVREVCNSG-----YKIEG 168
Db 154 WNETPICDRIPCGLPPTI-TNGDFISTNRE---NFIYGSVVTYRCNPGSGGRKRVFELVG 209
OY 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDVINGSPIS 204
Db 210 EPSIVCTSNDDQVGIVSGPAPQCIIPNKCTPPNVENGILVS 250
```

RESULT 6

```
LEM2_PIG ID LEM2_PIG STANDARD; PRT; 484 AA.
AC P98110;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE (CD62E).
GN SELE.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
```



```
RN RP SEQUENCE FROM N.A.
RX MEDLINE-92333688; PubMed-1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
RN [2]
RP SIMILARITY TO CCP.
RX MEDLINE-92260674; PubMed-1316492;
RA Albrecht J.-C., Fleckenstein B.;
RT "New member of the multigene family of complement control proteins in
herpesvirus saimiri.";
RL J. Virol. 66:3937-3940(1992).
CC -|- ALTERNATIVE PRODUCTS: A MEMBRANE-BOUND FORM AND A SECRETED FORM
CC ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC -|- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
CC COMPLEMENT ACTIVATION (RCA).
CC -|- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64346; CAA45626.1; -
DR EMBL; X64346; CAA45627.1; -
DR EMBL; X60283; CAA42823.1; -
DR EMBL; X60283; CAA42822.1; -
DR PIR; B42534; WMBE2E.
DR PIR; A42534; WMBE1E.
DR PIR; S24567; S24567.
DR HSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
KW Signal; Repeat; Sushi; Transmembrane; Alternative splicing;
KW Glycoprotein.
FT SIGNAL 1 20
FT CHAIN 21 360
FT DOMAIN 83 143
FT DOMAIN 146 206
FT DOMAIN 209 265
FT TRANSMEM 328 350
FT DISULFID 84 125
FT DISULFID 111 142
FT DISULFID 147 191
FT DISULFID 175 205
FT DISULFID 210 252
FT DISULFID 238 264
FT CARBOHYD 36 36
FT CARBOHYD 39 39
FT CARBOHYD 46 46
FT CARBOHYD 72 72
FT CARBOHYD 155 155
FT CARBOHYD 294 294
FT VARSPIC 289 302
FT VARSPIC 303 360
FT SEQUENCE 360 AA; 40006 MW; 6278A6C2ECD49669 CRC64;
Query Match 20.3%; Score 236; DB 1; Length 360;
Best Local Similarity 33.3%; Pred. No. 8e-14;
Matches 61; Conservative 17; Mismatches 89; Indels 16; Gaps 7;
QY 24 YPEGQAIYKCRPGYRSLGNVIMVCRKGEWALNPLRKCQKRCPCGPGDPTFGTFLTGG 83
DB 44 Y:NGTTLHTWCREGYAKRPVQITVCVNGWTV---PKCKKKKCSFPQDLLNGRTVT-G 99
```

```
QY 84 NVPEYGVKAVYTCNEGYYQLLGEIN----YRECDTDGWTNDIPICEVVKCLPVTAPENGKI 139
DB 100 LLS--YGSVITYTCNSCYSLIGSTTSACLLKRGGRVDMTPRPICDIKKCKPPQIANG-- 156
QY 140 VVSAMPDREYHFGQAVRFVCSNGYK--TEGDEEMHCSDGFW-SKEKPKCVSISKSPD 196
DB 157 --THTNVKDFYTYLDTVTYSCNDETKLTLTGSPSSKLCSETGSMVNPNGETKCEFIKCLKPQ 214
QY 197 VIN 199
DB 215 VAN 217
RESULT 13
DAFL_MOUSE
ID DAFL_MOUSE STANDARD; PRT; 390 AA.
AC Q61475; Q61397; P97732;
CT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Complement decay-accelerating factor, GPI-anchored precursor
DE (DAF-GPI).
GN DAF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-Testis;
RX MEDLINE-95403982; PubMed-7545711;
RA Spicer A.P., Seldin M.F., Gendler S.J.;
RT "Molecular cloning and chromosomal localization of the mouse decay-
RT accelerating factor genes. Duplicated genes encode
RT glycosylphosphatidylinositol-anchored and transmembrane forms.";
RL J. Immunol. 155:3079-3091(1995).
RN [2]
RP SEQUENCE OF 7-390 FROM N.A.
RC STRAIN-BALB/C; TISSUE-Spleen;
RX MEDLINE-96362213; PubMed-8671624;
RA Fukuoka Y., Yasui A., Okada N., Okada H.;
RT "Molecular cloning of murine decay accelerating factor by
RT immunoscreening.";
RL Int. Immunol. 8:379-385(1996).
CC -|- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY
CC SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -|- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE,
CC LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.
CC -|- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC -|- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC -|- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC (RCA) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L41366; AAB00091.1; -
DR EMBL; D63679; BAA09830.1; -
DR HSSP; P08603; IHCC.
DR MGD; MGI:104850; Daf1.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
KW Complement pathway; Glycoprotein; Membrane; Repeat; GPI-anchor;
```


Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	1163	100.0	1172	4	Q9NU87		Q9nu87 homo sapien
2	848	72.9	1236	11	Q91YB6		Q91yb6 homo sapien
3	499	42.9	669	6	Q28085		Q28085 bos taurus
4	368.5	31.7	1053	13	Q91275		Q91275 paratabrax
5	276.5	23.8	645	12	Q9NRU2		Q9wru2 macaca mula
6	275	23.6	360	12	Q9YQB8		Q9yqb8 ateline her
7	270.5	23.3	550	12	P88903		P88903 kaposi's sa
8	270.5	23.3	550	12	O40912		O40912 kaposi's sa
9	269.5	23.2	395	12	Q9J2M6		Q9j2m6 macaca mula
10	263.5	22.7	3389	4	Q960U9		Q960u9 homo sapien
11	263.5	22.7	3508	4	Q96RM4		Q96rm4 homo sapien
12	261.5	22.5	533	4	Q9H4W4		Q9h4w4 homo sapien
13	260	22.4	679	11	Q98254		Q98254 mus musculu
14	260	22.4	3567	11	Q9ES77		Q9es77 mus musculu
15	258.5	22.2	3564	11	Q923L3		Q923l3 mus musculu
16	257	22.1	347	6	Q9MYJ6		Q9myj6 pan troglod

```

Db      139  PICEVVKCLPVTAPENGKIVSSAMEPDREYHFQAVRFVCMNSYKIEGDEEMHCSDDGF 198
Qy      181  SKPKPCVCKSPDVINGSPISQKI 207
          |||||
Db      199  SKPKPCVCKSPDVINGSPISQKI 225

RESULT  2
Q911YB6 Q911YB6 PRELIMINARY; PRT; 1236 AA.
AC Q911YB6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE COMPLETE INHIBITORY FACTOR H.
GN FH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[]
RN RATTUS
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RA Demberg T., Goetze O., Schlaf G.;
RT "Rat complement factor H: molecular cloning, sequencing and expression
RT in tissues and isolated cells.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ320522; CAC67513.1; -.
SQ SEQUENCE 1236 AA; 140343 MW; 1AC89FFA28232EBF CRC64;

```

Query Match	72.9%	Score 848;	DB 11;	Length 1236;
Best Local Similarity	71.4%;	Pred. No. 6.9e-78;		
Matches 147;	Conservative 21;	Mismatches 38;	Indels 0;	Gaps 0;
Qy	1	EDCNELPRRTTEILLTGSWSBOTYEGTQATYVKCRPGYRSILGNVTWCRKGWVALNPLR	60	
Db	19	EDCKGPPPRENSEILLGSWSQLISECTQATYKCRPGYRTLGTIVKVCCKNGEWFPSFR	78	
Qy	61	KQKRPCGHPGDTTFGTTLTGNNVFYGVKAVYTCNKGYOLLGGINRECDTDCGWTNDI	120	
Db	79	ICRKRPCGHPGDTTFPGSFLAVGSEFEFGAKVVTYDCGYOLLGIDYRECDADGWTNDI	138	
Qy	121	PICEVVKCLPYTAPENGIVIVSSAMEPDREYHFGQAVRFFVCNMSGYKIEGDEEMHCSDGFGW	180	
Db	139	PICEVVKCLPYTELENGRIVISGAEPDOEYTFGQVVRFFECNSGFKIEGQEKEMHCSENGLW	198	
Qy	181	SKERPKCIVEISCKSPDVIINGSPISQK	206	
Db	199	SNKRPQCVEISCLPPRVENGDIYLVK	224	

RESULT	3	
Q28085		
ID	Q28085	PRELIMINARY; PRT: 669 AA.
AC	Q28085;	
DT	01-NOV-1996	(TREMBLrel. 01, Created)
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)
DE	CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).	
OS	Bos taurus	(Bovine).
OC	Eukaryota;	Metazoa;
OC	Mammalia;	Eutheria;
OC	Bovidae;	Bovinae;
OX	NCBI_TaxID=9913;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=LIVER;	
RA	MEDLINE=96202005; PubMed=8615824;	
RX	Soames C.J., Day A.J., Sim R.B.;	
RT	"Prediction from sequence comparisons of residues of factor H involved	
RL	in the interaction with complement component C3b.";	
RL	Biochem. J. 315:523-531(1996).	

```

DR EMBL; X98697; CAA67257.1; -.
DR HSP; P10998; 1VVD.
DR InterPro; IPR00436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 11.
DR SMART; SM00032; CCP; 11.
FT NON_TER 1
FT NON_TER 669
SQ SEQUENCE 669 AA; 75683 MW; D0D9DB30EE747AC2 CRC64;

Query Match      42.9%; Score 499; DB 6; Length 669;
Best Local Similarity 64.9%; Pred.No.1.6e-42;
Matches 85; Conservative 19; Mismatches 27; Indels 0; Gaps

Qy    76   GTFTLTGNNFYEGVKAVYTCNEGYYQLLGINTYRECDTDCGTNDIPICEVVKCLPVTPE 135
Db    3   GSPLAGNDFEYGAKVYVYTCDEGYQMVGEMNFRECNTNGWTNDIPICEVVKCLPVTPE 62

Qy    136  NGKTYSAMEPDREYHFQGAVRFVCNCSYRIEGDEEHMCHSDDGFWKSKEPKFCYEICKSP 195
Db    63  NGKIFSDALEPDQETTYGVQVVFECNSGYMLDGPQIHC SAGGWWSAETPKCVEIFCKPP 122

Qy    196  DVINGSPISQK 206
Db    123  VILNQAVLPK 133

```

```

RESULT      4
Q91275      PRELIMINARY;      PRRT; 1053 AA.
AC Q91275;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE COMPLEMENT REGULATORY PLASMA PROTEIN.
OS Parabrax nebulifer (barred sand bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Serranidae; Paralabrax.
OX NCBI_TaxID=30873;
RN [1]
RC TISSUE=LIVER;
RX MEDLINE=94318039; PubMed=8042982;
RA Dahmen A., Kaidoh T., Zipfel P.F., Gigli I.;
RT "Cloning and characterization of a cDNA representing a putative
RT complement-regulatory plasma protein from barred sand bass (Parablax
RT nebulifer).";
RL Blochem. J. 301:391-397(1994).
RL EMBL; L21703; AAA92556.1; -.
DR HSP; P08603; 1HFH.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 16.
DR SMART; SM00032; CCP; 16.
SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D5D3 CRC64;

```

	Query Match	31.7%	Score 368.5;	DB 13;	Length 1053;
	Best Local Similarity	39.6%;	Pred. No. 6.4e-29;		
	Matches	72;	Conservative 25;	Mismatches 76;	Indels 9; Gaps
QY	21	DOTPEGTQAIYKCRGYSLSGNVIMVKRGEWALNPLKKCKRCGHGDPDPFGTFTL	80	:	: :
Db	47	EASYPGRQRVRGNGVGS--GFELVCVEGWETRG--AKCQPSCGHGGDAQADFHL	102	:	: :
QY	81	TGGNVFEGYKAVYTCEGYQLLGEINRYRCDTDGWTNDIPICEYVKVLPTVPANGKIV	140	:	: :
Db	103	AEGNDFEFGSKVVYTCQKGQMYSRINYRCAVEAGDGVVPVCESQCPLIHVDNNQVI	162	:	: :
QY	141	SSAMEPDREYHFQGAVRFVCSNYKI--EGDEEMHCSDDGFSWKPKCVKEISCKSPDVIN	199	:	: :
Db	163	GG-----PEETAFGNVYRFSCKRSSEILDSGPFLCYDDERGWDGWPVKKAIFCALPPIEN	218	:	: :

QY 200 GS 201
 Db 219 GN 220

RESULT 5

Q9WRU2
 ID Q9WRU2 PRELIMINARY; PRT; 645 AA.
 AC Q9WRU2;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE COMPLEMENT BINDING PROTEIN.
 OS Macaca mulatta rhadinovirus 17577.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=83534;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99174001; PubMed=100741154;
 RA Searles R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;
 RT "Sequence and genomic analysis of a rhesus macaque rhadinovirus with
 RT similarity to Kaposi's sarcoma-associated Herpesvirus/Human
 RT herpesvirus 8";
 RL J. Virol. 73:3040-3053(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Searles R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF083501; AAD21332.1;
 DR HSSP: P10998; 1VVD.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 8.
 DR SMART: SM00032; CCP; 8.
 SQ SEQUENCE 645 AA; 71526 MW; 93D8DE35ABF61EB2 CRC64;

Query Match 23.8%; Score 276.5; DB 12; Length 645;
 Best Local Similarity 31.1%; Pred. No. 9.1e-20;
 Matches 65; Conservative 35; Mismatches 90; Indels 19; Gaps 9;
 QY 1 EDCNELPPRRNTEITLGSMSDOTYPEGTQAIYKCRPGYRSLGNVIMV--CRKGEWVALNP 58
 Db 21 ENCK--PPHFTYRVKSNTEKDLVSGETAELICRPGYVNTKIITTECLQNGTW--STP 76
 QY 59 LRKCKRCPCHGDPFGFTLTGG--NVFEYGVKAVYTCNEGQYLLGEINREC---DTD 114
 Db 77 NFPCKRCKPTPADLLNGAVHGGDNALKFGSNISYECNEGVDLIGS-NVRFCILQDTE 135
 QY 115 --GWTNDIPICEVVKCLPVTAPENKIVSSAMEPDRE-YHFGQAVRFVCSNGYKIEGDEE 171
 Db 136 NVNWSNEPVCBIQCKP2PAVEHGDYL-----PNQDVYNGDAITFKCSLSYTLVGSTT 190
 QY 172 MHCSDGDFW-SKEKPKVCEISCKSPDVING 200
 Db 191 LVCTSNKKNWSNFPCTCLMVCSQPIDNG 219

RESULT 6

Q9YTQ8
 ID Q9YTQ8 PRELIMINARY; PRT; 360 AA.
 AC Q9YTQ8;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE COMPLEMENT CONTROL PROTEIN HOMOLOG CCPH.
 OS Ateline herpesvirus 3.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=85618;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=73;
 RX MEDLINE=20091363; PubMed=10623770;
 RA Albrecht J.C.;
 RT "Primary structure of the Herpesvirus Ateles genome.";
 RL J. Virol. 74:1033-1037(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=73;
 RA Albrecht J.-C., Fleckenstein B.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF083424; AAC95530.1;
 DR HSSP: P10998; 1VVD.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 4.
 DR SMART: SM00032; CCP; 4.
 SQ SEQUENCE 360 AA; 40208 MW; 118CF83C034352A0 CRC64;

Query Match 23.6%; Score 275; DB 12; Length 360;
 Best Local Similarity 37.1%; Pred. No. 6.3e-20;
 Matches 78; Conservative 20; Mismatches 80; Indels 32; Gaps 14;
 QY 8 PRENTEI-----LTGSWSDOTYPEGTQAIYKCRPGYRSLGNVI--MVCRKGEWVALNPL 59
 Db 24 PKRNRVYSLRYVNITN--SGSYNPNGTTLQVTCRKG--IGROIQTVTCVNGNMTVPN-- 77
 QY 60 RKQCKRCPCHGDPFGFTLTGGNVFEYGVKAVYTCNEGQYLLGEINREC--DTDG-- 115
 Db 78 -EQCKRCSCTPADLLNGWYTVT--GNLY--VGSVITYTCNTGYLLGSPT--SCCLLGPDCRV 133
 QY 116 -WTNDIPICEVVKCLPVTAPENKIVSSAMEPDREYH-FQGAQVRFVCSNGYK--IEGDEE 171
 Db 134 NMTPRPPEICEITKCKPPTIANGHTNI-----KEYVTYLDVAVTSCNDETCLTLTGPS 188
 QY 172 MHCSDGDFW-SKEKPKVCEISCKSPDVING 200
 Db 189 KQCSGTGRWVDEETKCFKVKIPQVANG 218

RESULT 7

P88903
 ID P88903 PRELIMINARY; PRT; 550 AA.
 AC P88903;
 DT 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE ORF 4.
 OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=37296;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97094384; PubMed=8939871;
 RA Moore P.S., Boshoff C., Weiss R.A., Chang Y.;
 RT "Molecular mimicry of human cytokine and cytokine response pathway
 RT genes by KSHV";
 RL Science 274:1739-1744(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97121480; PubMed=8962146;
 RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
 RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
 RT "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
 RT (HHV8)";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
 RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U75698; AAC57082.1;
 DR HSSP: P10998; 1VVD.

```
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 4.
DR SMART: SM00032; CCP; 4.
SQ SEQUENCE 550 AA; 60688 MW; D4B8B2B4BACD1CB5 CRC64;

Query Match
Best Local Similarity 23.3%; Score 270.5; DB 12; Length 550;
Matches 59; Conservative 24; Mismatches 80; Indels 17; Gaps 7;

QY 33 KCRPGYRSLG-NVIMVC-RKGEWALNPLRKCCKRCGHPGCDTPFGTFTLTGG-NVFEYG 89
Db 52 RCRSGYTYARNITATCLQGWT--SEPTATCNKSKCPNPFGEIQNGKVIHFHGGQDALKYG 109
QY 90 VKAVYTCNEGYQLLGEINREC-----DTGWTNDIPICEVVKCLPVTAPENCKIVSSAM 144
Db 110 ANISYVCNEGYFLVGREYRYCMIGASQGMWSSPPFCEKEC-----HRPKIENGDF 163
QY 145 EPDREYH-FQGAQVRFVNCYSKYIEGDEMHCSDDGFWSEKPKCKVEISCKSPDVINGSPI 203
Db 164 KPDQYIEYNDVAHFECNEGYTLVGPHSIACAVNNTWTSNMPTCELACGCKFPSPVTHGYPI 223

RESULT 8
ID O40912 PRELIMINARY; PRT; 550 AA.
AC O40912;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF 04.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=37296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97296220; PubMed=9151804;
RA Neipel F., Albrecht J.C., Fleckenstein B.;
RT "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus
RT human herpesvirus 8: determinants of its pathogenicity?";
RL J. Virol. 71:4187-4192(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Neipel F., Albrecht J.-C., Ensser A., Huang Y.-Q., Li J.J.,
RA Friedman-Kien A.E., Fleckenstein B.;
RT "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: U93872; AAB82602.1; -.
DR HSSP: P10998; IVD.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 4.
DR SMART: SM00032; CCP; 4.
SQ SEQUENCE 550 AA; 60648 MW; 3A7FC1C30F79C6A6 CRC64;

Query Match
Best Local Similarity 23.3%; Score 270.5; DB 12; Length 550;
Matches 59; Conservative 24; Mismatches 80; Indels 17; Gaps 7;

QY 33 KCRPGYRSLG-NVIMVC-RKGEWALNPLRKCCKRCGHPGCDTPFGTFTLTGG-NVFEYG 89
Db 52 RCRSGYTYARNITATCLQGWT--SEPTATCNKSKCPNPFGEIQNGKVIHFHGGQDALKYG 109
QY 90 VKAVYTCNEGYQLLGEINREC-----DTGWTNDIPICEVVKCLPVTAPENCKIVSSAM 144
Db 110 ANISYVCNEGYFLVGREYRYCMIGASQGMWSSPPFCEKEC-----HRPKIENGDF 163
QY 145 EPDREYH-FQGAQVRFVNCYSKYIEGDEMHCSDDGFWSEKPKCKVEISCKSPDVINGSPI 203
Db 164 KPDQYIEYNDVAHFECNEGYTLVGPHSIACAVNNTWTSNMPTCELACGCKFPSPVTHGYPI 223

RESULT 9
ID O9J2M6 PRELIMINARY; PRT; 395 AA.
AC O9J2M6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE COMPLEMENT BINDING PROTEIN.
OS Macaca mulatta rhadinovirus 26-95.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=119193;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20173730; PubMed=10708456;
RA Alexander L., Denekamp L., Knapp A., Auerbach M.R., Damania B.,
RA Desrosiers R.C.;
RT "The primary sequence of rhesus monkey rhadinovirus isolate 26-95:
RT sequence similarities to Kaposi's sarcoma-associated herpesvirus and
RT rhesus monkey rhadinovirus isolate 17577.";
RL J. Virol. 74:3388-3398(2000).
DR EMBL: AF210726; AAF59982.1; -.
DR HSSP: P10998; IVD.
DR InterPro: IPR001230; Prenyltn.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 4.
DR SMART: SM00032; CCP; 4.
DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1.
SQ SEQUENCE 395 AA; 43922 MW; B4C9C6F2E226AE06 CRC64;

Query Match
Best Local Similarity 23.2%; Score 269.5; DB 12; Length 395;
Matches 65; Conservative 34; Mismatches 87; Indels 19; Gaps 10;

QY 7 PPRNTEILTGSMSDQTYPEGTQAIYKCRPG-YRSLGNVIMVC-RKGEWALNPLRKCCK 64
Db 30 PPEDRFVMTKAN-QENYAVGTRVELICRPGFYKQANVYVECLSNGTWT--TPNAECRR 86
QY 65 RCGHPCDTPFGTFTLT-GGVNFEYGVKAVTTCNEGYQLLGEINREC-----DTD--GWT 117
Db 87 KRCSNPEDILNGEVIITDSNNAFKGNSNITYKNTGTVLLLG-ATVTRCTLLKYSNLVDWQ 145
QY 118 NDIPICEVVKCLPVTAPENCKIVSSAMEPDRE-YHFGQAVRFVNCYSKYIEGDEMHCS 176
Db 146 PAAPTCEIECKKQPDIECKYY-----PVQEPYNYLETIFTCKNDFSLIGNTTTMT 200
QY 177 DGFWSKEKPKCKVEISCKSPDVINGS 201
Db 201 NGTWSSPVKPKQITCSAPNIDHGT 225

RESULT 10
ID Q96Q09 PRELIMINARY; PRT; 3389 AA.
AC Q96Q09;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CUB AND SUSHI MULTIPLE DOMAINS PROTEIN 1 SHORT FORM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21365705; PubMed=11472063;
RA Sun P.C., Uppaluri R., Schmidt A.P., Pashia M.E., Quant E.C.,
RA Sunwoo J.B., Gollin S.M., Scholnick S.B.;
RT "Transcript map of the 8p23 putative tumor suppressor region.";
RL Genomics 75:17-25(2001).
DR EMBL: AY017307; AAG52948.1; -.

```

```
SQ SEQUENCE 3389 AA; 370293 MW; 53C3009FCD3ED76D CRC64;
Query Match 22.7%; Score 263.5; DB 4; Length 3389;
Best Local Similarity 29.6%; Pred. No. 1.5e-17;
Matches 55; Conservative 33; Mismatches 89; Indels 9; Gaps 6;
QY 14 ILTGWSDDQTYPECTQAIYKRCGRYSLGNVIMVC-RKGEWVALNPLRKCKRCPGHPGD 72
D 2630 IVNGHISGDGFSYRDVTVVYQCNGFRLVGTSVRICLDQHKWSGQTPV--CVPITCGHPGN 2687
QY 73 TPGTFTLTGNNVFEYGVKAVYTCNEGQYLLGEINRYECDDTG-WTNDIPICEVVKCLPV 131
D 2688 PAHG---FTNGSEFNLDVNVFTCNTGYLLQG-VSRAQCNSNGQWSSPLPTCRVNVCSDP 2743
QY 132 TAPENKIVSSAMEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFSWKEPKVCVEIS 191
D 2744 GFVENA-IRHQGNFPESFEYGNISLYHCKKGFHLLGSSALTCMANGLDWRLPKCLAIS 2802
QY 192 CKSPDV 197
D 2803 CGHPGV 2808
RESULT 11
Q96RM4 PRELIMINARY; PRT; 3508 AA.
AC Q96RM4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CUB AND SUSHI MULTIPLE DOMAINS 1 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21365705; PubMed=11472063;
RA Sun P.C., Uppaluri R., Schmidt A.P., Pashia M.E., Quant E.C.,
RA Sunwoo J.B., Gollin S.M., Scholnick S.B.;
RT "Transcript map of the 8p23 putative tumor suppressor region.";
RL Genomics 75:17-25(2001).
DR EMBL; AF333704; AAK73475.1; -.
SQ SEQUENCE 3508 AA; 382824 MW; 9268C3EBF3F78C18 CRC64;
Query Match 22.7%; Score 263.5; DB 4; Length 3508;
Best Local Similarity 29.6%; Pred. No. 1.5e-17;
Matches 55; Conservative 33; Mismatches 89; Indels 9; Gaps 6;
QY 14 ILTGWSDDQTYPECTQAIYKRCGRYSLGNVIMVC-RKGEWVALNPLRKCKRCPGHPGD 72
D 2630 IVNGHISGDGFSYRDVTVVYQCNGFRLVGTSVRICLDQHKWSGQTPV--CVPITCGHPGN 2687
QY 73 TPGTFTLTGNNVFEYGVKAVYTCNEGQYLLGEINRYECDDTG-WTNDIPICEVVKCLPV 131
D 2688 PAHG---FTNGSEFNLDVNVFTCNTGYLLQG-VSRAQCNSNGQWSSPLPTCRVNVCSDP 2743
QY 132 TAPENKIVSSAMEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFSWKEPKVCVEIS 191
D 2744 GFVENA-IRHQGNFPESFEYGNISLYHCKKGFHLLGSSALTCMANGLDWRLPKCLAIS 2802
QY 192 CKSPDV 197
D 2803 CGHPGV 2808
RESULT 12
Q9H4W4 PRELIMINARY; PRT; 533 AA.
AC Q9H4W4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
```

```
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DJ947L8.1.8 (NOVEL SUSHI (SCR REPEAT) DOMAIN PROTEIN) (FRAGMENT).
GN DJ947L8.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wallis J.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; ALJ55178; CAC10283.1; -.
DR HSPB; P10998; 1VVD.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sush1; 8.
DR SMART; SM00032; CCP; 8.
FT NON_TER 1
FT NON_TER 533
SQ SEQUENCE 533 AA; 57921 MW; 0554F7E10911F9BF CRC64;
Query Match 22.5%; Score 261.5; DB 4; Length 533;
Best Local Similarity 30.4%; Pred. No. 2.5e-18;
Matches 58; Conservative 29; Mismatches 91; Indels 13; Gaps 7;
QY 14 ILTGWSDDQTYPECTQAIYKRCGRYSLGNVIMVC-RKG-EMVALNPLRKCKRCPGHPGD 72
D 306 IVNGHINGENYSYRGVYQCNGFRLIGHMSVRICQDQHHWSGKTPF--CVPITCGHPGN 363
QY 73 TPGTFTLTGNNVFEYGVKAVYTCNEGQYLLGEINRYECDDTG-WTNDIPICEVVKCLPV 131
D 364 PVANG---LTGQNFNLNDVYKFCNPGYMAEGAAR-SOCLASGOWSDMLPTCRIINCTDP 419
QY 132 TAPENKIVSSAMEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFSWKEPKVCVEIS 191
D 420 GHQENSVRQVHSGPHR-FSGGTVSYRCNHRGFYLLGTPVLSCQGDGTWDRPQCCLVS 478
QY 192 CKSPDVINGSP 202
D 479 CGHP---GSP 485
RESULT 13
Q99254 PRELIMINARY; PRT; 679 AA.
AC Q99254;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C.
RX MEDLINE=90229754; PubMed=2139460;
RA Kurtz C.B., O'Toole E., Christensen S.M., Weis J.H.;
RT "The murine complement receptor gene family. IV. Alternative splicing
of Cr2 gene transcripts predicts two distinct gene products that share
homologous domains with both human CR2 and CR1."
RL J. Immunol. 144:3581-3591(1990).
RN [2]
RP SEQUENCE OF 21-367 FROM N.A.
RX MEDLINE=95105691; PubMed=7528766;
RA Kim Y.U., Kinoshita T., Molina H., Hourcade D., Seya T., Wagner L.M.,
RA Holers V.M.;
RT "Mouse complement regulatory protein Cr2/p65 uses the specific
mechanisms of both human decay-accelerating factor and membrane
cofactor protein.";
```


Query Match	22.4%;	Score 260;	DB 11;	Length 3567;
Best Local Similarity	29.1%;	Pred. No. 3.6e-17;		
Matches 55;	Conservative 30;	Mismatches 92;	Indels 12;	Gaps 4;
QY 17 GSWSDQTYPBCTQAIYACRPGYRSLGNVIMVC-RKGEWALNPLRKQKRCGHPGDTPF 75	1797 GHSSGEIYTVGTAVTFSCDEGHVGVSTICTLETGEMDLRP--SCEAISCQVP---PV 1851			
QY 76 GTFTLGCNNVFEGVGVKAVYTCNEGYYQLLGINTRECOTDGTNDIPICEVVKCLPVTAP 135	1852 PEGGVGDGSFTYKGYVYRCDYKGYTLTSGDEESACLAGSWSHSSPYCELVKCSQPEDIN 1911			
QY 136 NGKIVSSAMEPDREHYHFGQAVRVCNGSYGKIEDEEMHCSDDGFWSKKPKVCVEISCKSP 195	1912 NGXYILSLGT-----YLSIASYSCNGSYLQGPSLLETCTAGSWDRAPPSQQLVSCGEP 1965			
QY 196 DVINGSPIS 204	1966 PIVKDAVIT 1974			
RESULT 15				
Q923L3	PRELIMINARY;	PRT: 3564 AA.		
ID Q923L3				
AC Q923L3				
DT 01-DBRC-2001 (TrEMBLrel. 19, Created)				
DT 01-DBRC-2001 (TrEMBLrel. 19, Last sequence update)				
DE 01-DBRC-2001 (TrEMBLrel. 19, Last annotation update)				
DE CSMD1				
GN CSMD1				
OS Mus musculus (Mouse).				
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
NCBI_TaxID=10090;				
RN [1]				
RP SEQUENCE FROM N.A.				
RC STRAIN=C57BL/6;				
RA MEDLINE=21365705; PubMed=11472063;				
RA Sun P.C., Uppaluri R., Schmidt A.P., Pashia M.E., Quant E.C.,				
RA Sunwoo J.B., Gollin S.M., Scholnick S.B.;				
RT "Transcript map of the gp23 putative tumor suppressor region.";				
RL Genomics 75;17-25(2001)				
RL EMBL: AY01475; AAG54083.1; --				
SR SEQUENCE 3564 AA; 387865 MW; 70824C55B0674609 CRC64;				
QY 1797 GHSSGEIYTVGTAVTFSCDEGHVGVSTICTLETGEMDLRP--SCEAISCQVP---PV 1851				
QY 76 GTFTLGCNNVFEGVGVKAVYTCNEGYYQLLGINTRECOTDGTNDIPICEVVKCLPVTAP 135				
QY 136 NGKIVSSAMEPDREHYHFGQAVRVCNGSYGKIEDEEMHCSDDGFWSKKPKVCVEISCKSP 195				
QY 196 DVINGSPIS 204				
QY 1966 PIVKDAVIT 1974				

QY 14 ILTGSWSDOTYPEGTQAIYKCRPGYRSLGNVIMVC-----RKGEWVALNPLRKCQKRPCG 68
Db 2686 IVNGHISGDGFSYRDTVVYQCNPGRVLGTSVRICRTTSGRGR-----LTVCVPITCG 2739
QY 69 HPGDTFFGTFTLTGGNVFEYGVKAVYTCNEGYOLLGEINYRECDTDG-WTNDIPICEVVK 127
Db 2740 HPGNPAHG---LTNGTEFNLDLVNFTCHTGYRLQG-ASRAQCRSNGQWSSPLPICRVVN 2795
QY 128 CLPVTAPENCKIVSSAMEPDREYHFGQAVFVCNCGYKIEGDEEMHCSDDGFWSKKPKC 187
Db 2796 CSDPGSVENAVRHGQONFPE-SFEYGT\$VWYHCKTGFLYLLGSSALTCMASGLWDRSLPKC 2854
QY 188 VEISCKSPDV 197
Db 2855 LAISCGHPGV 2864

Search completed: August 29, 2002, 15:11:38
Job time: 599 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:00:49 ; Search time 75.55 Seconds
(without alignments)
304.332 Million cell updates/sec

Title: US-09-316-163-9

Perfect score: 1163

Sequence: 1 EDCELPPRNTTEILGWS.....VEISCKSPDVINGSPISOKI 207

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	721	62.0	240	18 AAW39154	Human partial Comp
2	627	53.9	216	18 AAW39155	Clone PRB9PH410 C
3	276.5	23.8	645	21 AAB53125	Macaca mulatta rha
4	261.5	22.5	613	22 AAU00816	Human Immunoglobul
5	261.5	22.5	882	22 AAB83372	NOV16 protein sequ
6	261	22.4	299	17 AAW06881	Decay accelerating
7	261	22.4	376	20 AAY50035	Human complement r
8	261	22.4	381	8 AAP70048	Human decay accele
9	261	22.4	381	10 AAP94773	Decay accelerating
10	261	22.4	381	16 AAR66683	Decay accelerating
11	261	22.4	381	18 AAW26317	Human decay accele

12	261	22.4	381	20 AAY31740	Human CD55 and 791
13	261	22.4	381	20 AAW73505	Decay accelerating
14	261	22.4	440	8 AAP70049	Human decay accele
15	261	22.4	440	10 AAP94774	Membrane bound dec
16	261	22.4	440	16 AAR66684	Decay accelerating
17	261	22.4	440	18 AAW27483	Human glycophospha
18	261	22.4	543	13 AAR28557	CRI-4 (99H, 103E)
19	261	22.4	577	17 AAW06882	Membrane co-factor
20	261	22.4	611	22 AAE12569	CAB2 protein. Uni
21	261	22.4	611	22 AAE03762	CAB-2 chimeric pro
22	259	22.3	497	22 ABB10508	Human CDNA SEQ ID
23	259	22.3	497	22 AAU18149	Novel human uterin
24	259	22.3	497	22 AAU17035	Human novel secret
25	259	22.3	497	22 AAU19960	Novel human calciu
26	259	22.3	515	22 AAM93953	Human polypeptide,
27	259	22.3	543	13 AAR28547	CRI-4 (52S, 53S, 5
28	259	22.3	1139	22 ABB10326	Human CDNA SEQ ID
29	259	22.3	1139	22 AAU18126	Novel human uterin
30	259	22.3	1139	22 AAU16963	Human novel secret
31	259	22.3	1139	22 AAU19902	Novel human calciu
32	257.5	22.1	496	20 AAY55752	Human CRI protein
33	257.5	22.1	543	13 AAR28560	CRI-4 (114S) analo
34	256.5	22.1	263	22 AAB48846	Cowpox virus infla
35	255.5	22.0	320	22 AAG68150	Codon modified hum
36	255	21.9	254	15 AAR47154	Sequence of solubi
37	255	21.9	254	15 AAR47155	Sequence of solubi
38	255	21.9	543	13 AAR28550	CRI-4 (64K) analog
39	255	21.9	543	13 AAR28553	CRI-4 (85R, 87N) a
40	255	21.9	543	13 AAR28565	CRI-4 (121Q) analo
41	255	21.9	543	13 AAR28566	CRI-4 (318R, 319N)
42	255	21.9	543	13 AAR28567	CRI-4 (318-321 RNP
43	255	21.9	543	13 AAR28568	CRI-4 (347T, 349Y)
44	255	21.9	543	13 AAR28569	CRI-4 (369-376 STK
45	255	21.9	543	13 AAR28570	CRI-4 (266-274 CLK

ALIGNMENTS

RESULT 1

AAW39154	ID AAW39154 standard; Protein; 240 AA.
XX	AC AAW39154;
XX	DT 27-APR-1998 (first entry)
XX	DE Human partial Complement factor H protein fragment 1.
XX	DE Complement factor H; tumour associated antigen; renal cancer;
XX	KW urogenital cancer; medicament; modulator.
XX	OS Homo sapiens.
XX	PN WO9738136-A1.
XX	PD 16-OCT-1997.
XX	PF 09-APR-1997; 97WO-US05710.
XX	PR 06-MAR-1997; 97US-0812481.
PR	09-APR-1996; 96US-0015083.
PR	09-APR-1996; 96US-0630048.
PR	06-MAR-1997; 97US-0038614.
PA	(BARD-) BARD DIAGNOSTIC SCI INC.
PI	Enfield DL, Hass GM, Kinders RJ;
DR	WPI; 1997-512742/47.
XX	N-PSDB; AAV02790.
PT	Treating or screening for cancer, e.g. renal or urogenital cancer -

PT by modulating or detecting tumour associated human complement Factor
PT H related antigen, or nucleic acid encoding it
XX
PS Example 6B; Fig 6B; 104pp; English.
XX
CC This partial protein sequence represents a region of the human
CC tumour-associated complement factor H (CFH). This sequence is used
CC in the identification of complement factor H related proteins and
CC antigens isolated from clone pRRB9FH410 (see AAW39155). The detection of
CC such proteins and a CFH antigens can be used in screening or for the
CC treatment of renal or urogenital cancer, e.g. bladder, cervical or
CC prostate cancer. Agents that may modulate this antigen could be used in
CC the manufacture of a medicament for the treatment of a tumour cell.
XX
SQ Sequence 240 AA;

Query Match 62.0%; Score 721; DB 18; Length 240;
Best Local Similarity 100.0%; Pred. No. 3.5e-59;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 FTLTGNVFEGYGVKAVTCTNEGQYLLGEINRYRECDDTGDWTDIPICEVVKCLPVTAPENG 137
|||||
Db 1 ftiltgnvfeygvkavtctnegyqllgeinryrecddtgdwtndipicevkvclpvtapeng 60

QY 138 KIVSSAMEPDREYHFGQAVRVCNSGYKIEGDEEMHCDDGFWSKPKCIVEISCKSPDV 197
|||||
Db 61 kivssamepdreyhfgqavrfvcnsgykiegdeemhcsddgfwskpkcveisckspdv 120

QY 198 INGSPTISQKI 207
|||||
Db 121 ingspisqki 130

RESULT 2
AAK39155
ID AAW39155 standard; Protein; 216 AA.
XX
AC AAW39155;
XX
DT 27-APR-1998 (first entry)
XX
DE Clone pRRB9FH410 CFH related protein fragment.
XX
KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator.
XX
OS Synthetic.
XX
PN WO9738136-A1.
XX
PD 16-OCT-1997.
XX
PF 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
PR 09-APR-1996; 96US-0015083.
PR 09-APR-1996; 96US-0630048.
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
PI Enfield DL, Hass GM, Kinders RJ;
XX
DR WPI; 1997-512742/47.
DR N-PSDB; AAV02791.
XX
XX Treating or screening for cancer, e.g. renal or urogenital cancer -
PT by modulating or detecting tumour associated human complement factor -
PT H related antigen, or nucleic acid encoding it
XX
PS Example 6B; Fig 6B; 104pp; English.
XX

CC This partial protein is found in clone pRRB9FH410 and represents a
CC complement factor H related protein with homology to a region of the
CC human tumour-associated complement factor H (CFH). The detection of this
CC protein and a CFH antigen can be used in screening or for the treatment
CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
CC Agents that may modulate this antigen could be used in the manufacture of
CC a medicament for the treatment of a tumour cell.
XX
SQ Sequence 216 AA;

Query Match 53.9%; Score 627; DB 18; Length 216;
Best Local Similarity 99.1%; Pred. No. 1.7e-50;
Matches 112; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 95 TCNEGYQLLGEINRYRECDDTGDWTDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQ 154
|||||
Db 1 tcnegyqllgeinryrecddtgdwtndipicevkvclpvtapengkivssamepdreyhfgq 60

QY 155 AVRFVCNSGYKIEGDEEMHCDDGFWSKPKCIVEISCKSPDVINGSPISQKI 207
|||||
Db 61 avrfvcnsgykiegdeemhcsddgfwgkpkcveisckspdvingspisqki 113

RESULT 3
AAB53125
ID AAB53125 standard; Protein; 645 AA.
XX
AC AAB53125;
XX
DT 28-FEB-2001 (first entry)
XX
DE Macaca mulatta rhadinovirus 17577 RRV ORF4 protein SEQ ID NO:7.
XX
KW Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
KW genome; Kaposi's sarcoma-associated herpesvirus; KSHV; Interleukin 6;
KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
KW cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KW splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.
XX
OS Macaca mulatta rhadinovirus 17577.
XX
PN WO200028040-A2.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26260.
XX
PR 06-NOV-1998; 98US-0107507.
PR 20-NOV-1998; 98US-0109409.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Wong SW, Axthelm MK, Searles RP;
XX
DR WPI; 2000-376552/32.
XX
PT New rhesus rhadino virus for producing non-human primate model useful
PT for testing potential treatments and efficacy of the candidate vaccine
PT for conditions associated with RRV infection -
XX
PS Claim 5; Page 122-123; 141pp; English.
XX
CC The present invention describes a novel rhesus macaque rhadinovirus
CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
CC RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
CC encoded by the genome sequence. The present invention also specifically
CC claims the individual open reading frame (ORF) nucleotide sequences from
CC the genome which encode the individual proteins, but these sequences are
CC not given. A non-human animal infected with RRV can be used for testing
CC the efficacy of drug in the treatment of condition associated with
CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative

[illegible]

QY 73 TPFGTFLTCGNVFEXGVKAVYTCNEGQYLLGEINRECDTDG-WTNDIPICEVVKCLPV 131
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
 Db 275 pvnq---ltgngfnIndvkvfcnpymaegaar-sqcllasgqsdmlptcrliinctdp 330
 QY 132 TAPENKIVSSAMEPDREYHFGQAVRVCNKGKIEGDEMHCSDDGFWSEKPKVCVEIS 191
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
 Db 331 ghqensvrqvhagpphr-fsfgttvsyrcnhgfyllgtplscqgdgtwdrprpqcllvs 389
 QY 192 CKSPDVINGSP 202
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
 Db 390 cghp-----gsp 396
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
 RESULT 5
 AAB83372
 ID AAB83372 standard; Protein; 882 AA.
 AC AAB83372;
 DT 26-MAR-2002 (first entry)
 XX
 DE NOV16 protein sequence.
 XX
 KW NOV; Cytostatic; contraceptive; antiinflammatory; immunomodulatory;
 KW cardiovascular; casein kinase II phosphorylation site; contraception;
 KW serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;
 KW epidermal growth factor; cell development; apoptosis; cell adhesion;
 KW growth migration; cell structure; motility; cancer; immune disorder;
 KW inflammatory disorder; cellular adhesion disorder; long-QT syndrome;
 KW cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;
 KW therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;
 KW NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.
 XX
 OS Unidentified.
 OS
 PN WO200136638-A2.
 XX
 PD 25-MAY-2001.
 XX
 PF 17-NOV-2000; 2000WO-US31543.
 XX
 PR 19-NOV-1999; 99US-166336P.
 PR 29-NOV-1999; 99US-167785P.
 PR 08-MAR-2000; 2000US-187844P.
 PR 16-NOV-2000; 2000US-071541P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Lichenstein H, Vernet C, Fernandes E;
 XX
 XX WPI; 2001-648134/74.
 DR N-PSDB; AAF87127.
 XX
 PT Novel human polypeptides and the nucleic acids that encode them useful
 PT for preventing, diagnosing and treating e.g. cancer, inflammation and
 PT immune disorders -
 XX
 PS Claim 1; Page 50-52; 141pp; English.
 XX
 CC This sequence is the NOV16 protein. The invention relates to
 CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have
 CC Cytostatic; contraceptive; antiinflammatory; immunomodulatory; and
 CC cardiovascular activities. The sequences may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate NOVX
 CC expression. They may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of protein by expressing inactive proteins or to
 CC supplement the patients' own production of protein. They are used to
 CC produce NOVX proteins, by inserting the nucleic acid into a cell and
 CC culturing it to express the protein. The DNA may be used as DNA probes in
 CC assays to detect and quantitate the presence of similar DNAs in samples,
 CC and which patients may need restorative therapy. The NOVX protein may
 CC also be used as antigens in the production of antibodies (Abs) against

CC NOVX and in assays to identify modulators of NOVX expression and
 CC activity. The anti-NOVX Abs and antagonist are used to down regulate
 CC expression and activity. The anti-NOVX Abs are used for detecting the
 CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
 CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,
 CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
 CC characteristic of serine/threonine kinases, and are used to treat
 CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
 CC proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to
 CC the epidermal growth factor (EGF)-like super family and are involved in,
 CC e.g. regulation of cell development, apoptosis, cell adhesion, growth
 CC migration, cell structure and motility and protein management, and are
 CC used to treat cancers, inflammatory disorders, immune disorders and
 CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like
 CC fibrillin proteins and are used to treat cardiovascular disease e.g.
 CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
 XX
 XX SQ Sequence 882 AA;
 Query Match 22.5%; Score 261.5; DB 22; Length 882;
 Best Local Similarity 30.4%; Pred. No. 1e-15;
 Matches 58; Conservative 29; Mismatches 91; Indels 13; Gaps 7;
 QY 14 ILTGSMSDQTYPEGTQAIYKCRPGYSLGNVIMVCRKG-EWVALNPLRKCQKRPCGHPGD 72
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
 Db 217 ivnghingenysyrgsvyqcnafrlignsvricqgdhwsqktpf--cypitcghpgn 274
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
 QY 73 TPFGTFLTCGNVFEXGVKAVYTCNEGQYLLGEINRECDTDG-WTNDIPICEVVKCLPV 131
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
 Db 275 pvnq---ltgngfnIndvkvfcnpymaegaar-sqcllasgqsdmlptcrliinctdp 330
 QY 132 TAPENKIVSSAMEPDREYHFGQAVRVCNKGKIEGDEMHCSDDGFWSEKPKVCVEIS 191
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
 Db 331 ghqensvrqvhagpphr-fsfgttvsyrcnhgfyllgtplscqgdgtwdrprpqcllvs 389
 QY 192 CKSPDVINGSP 202
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
 Db 390 cghp-----gsp 396
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
 RESULT 6
 AAW06881
 ID AAW06881 standard; Protein; 299 AA.
 XX
 AC AAW06881;
 XX
 DT 18-MAR-1997 (first entry)
 XX
 DE Decay accelerating factor.
 XX
 KW Complement inhibitor; membrane co-factor protein; MCP;
 KW decay accelerating factor; DAF; chimeric protein; glycosaminoglycan;
 KW heparin; cell lysis; sepsis; adult respiratory distress syndrome;
 KW reperfusion injury; cell damage.
 XX
 OS Homo sapiens.
 XX
 PN WO9634965-A2.
 XX
 PD 07-NOV-1996.
 XX
 PF 03-MAY-1996; 96WO-US06301.
 XX
 PR 05-MAY-1995; 95US-0435149.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Creasey AA, Innis MA, Zaror I;
 XX
 DR WPI; 1996-506167/50.
 XX
 PT Chimeric proteins for inhibiting complement-mediated cell lysis -

comprise membrane co-factor protein and decay accelerating factor peptide sequences

Disclosure; Page 25; 33pp; English.

A portion (AAW06881) of the complement-inhibitor, decay accelerating factor (DAF), is used in novel chimeric proteins of formula A-R1-B-R2-C, where A and C are peptides (AAW06875-79, AAW06883-90) able to bind glycosaminoglycans (esp. heparin) present on cell surfaces, R1 is a portion of DAF or membrane co-factor protein (MCP, see also AAW06880), R2 is DAF when R1 is MCP or MCP when R1 is DAF, and B is a peptide that may have complement inhibitor activity. The chimeric proteins (see also AAW06882) are directed to cell surfaces where they inhibit complement-mediated cell lysis. They are used to treat and prevent disease states in which complement plays a role, e.g. sepsis, adult respiratory distress syndrome, reperfusion injury and tissue damage.

Sequence 299 AA;

Query Match 22.4%; Score 261; DB 17; Length 299;
Best Local Similarity 29.9%; Pred. No. 3.1e-16;
Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

Qy 1 EDCN---ELPRRNTEILTGSWSDQY-PEGTAIYKCRPGYR---SLGNVIMVCRKGEW 53
||| ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| : : :
Db 34 efcncsrcevptrlnsasikqpyitqnyfpvgtvveyecrpgyrrepslspklctclqnlkw 93
||| ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| : : :
Qy 54 VALNPLRKCQRPCCHPGDTFPGFTLTGGNVFEYGVKAVYTCNBYGOLLGEINYPRECDT 113
Db 94 --stavefckkkscnpngeirngldvpggllf--gatisfscntgyklfgsts-sfcll 148
||| ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| : : :
Qy 114 DG----WTNDIPICEVVKCLPVTAPENKIVSSAMEPDRE-VHFGQAVRFVNCNGYKIEG 168
||| ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| : : :
Db 149 sgssqvswdpipeireiyc-----pappqindgiqgerdhgygrqsvyackngftmlg 203
||| ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| : : :
Qy 169 DEEMHC-----SDGFWFSKSKPKC-----VEISCKSPDVIN-----GSPISQK 206
||| ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| : : :
Db 204 ehslyctvnndegewsgpppcegrgksltskypptvqkpttnvnppttevsptsqk 257
||| ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| : : :
RESULT 7
AAI50035
ID AAY50035 standard; Protein; 376 AA.
XX AC AAY50035;
XX DT 19-JAN-2000 (first entry)
XX DE Human complement regulatory protein DAF.
XX DE
XX KW Complement regulatory protein; decay accelerating factor; DAF;
XX KW immune reaction; xenotransplantation; xenograft; transplant; organ;
XX KW rejection; hyperacute; inhibition; protection; heart; lung; liver;
XX KW kidney; pancreas; thyroid; islet cell; neurons; stem cell; tissue;
XX KW skin.
XX OS Homo sapiens.
XX DE
XX DE
XX PN W09953042-A2.
XX XX
XX PD 21-OCT-1999.
XX XX
XX XX 08-APR-1999; 99WO-GB01085.
XX XX
XX PR 09-APR-1998; 98GB-0007520.
XX XX
XX XX (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
XX XX
XX PI Morgan BP, Rushmere NK, Hinchliffe SJ, Van Den Berg CW;
XX DR WPI; 1999-620420/53.

PA (GETH) GENENTECH INC.
 XX Caras IW;
 XX WPI; 1987-308481/44.
 DR N-PSDB; AAN70047.
 XX
 PT New decay accelerating factor variants - obtained with the factor by
 PT using recombinant DNA procedures.
 XX
 XX Disclosure; Page 15-17; 20pp; English.
 XX
 CC The probable phosphatidylinositol derivatization site is Cys(330).
 CC The DAF variant is useful for treating paroxysmal nocturnal
 CC haemoglobinuria, or inflammatory or cell lytic autoimmune
 CC diseases. It may be used to ameliorate allograft rejection
 CC or autoimmune diseases. See also AAN70046, AAN70048.
 XX
 SQ Sequence 381 AA;

Query Match 22.4%; Score 261; DB 8; Length 381;
 Best Local Similarity 29.9%; Pred. No. 4.1e-16;
 Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

QY 1 EDCN---ELPPRRNTEILTGSHSDQTY-PEGTQAIYKCRPGYR---SLGNVIMVCRKGEW 53
 Db | | | | | : | : | | | | | | | | | | : | : |
 92 efncrscevpnlrnsaslkqpyitqnyfpvgtvveecrpgyrrepslsplktclqnlkw 151
 QY 54 VALNPLRKCOKRCPGHPGDTPEFTLTGNGVFEYGVKAVYTCNEGYQLLGEINRECDT 113
 Db : | | : | | | : | : | | | : | : | | | | : | : |
 152 --stavefckkscnpngeirngqidvpqglf--gatisfscntgyklfgsts-sfc11 206

QY 114 DG----WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRFVCSGYKIEG 168
 Db | | | : | | : | : | : | : | : | : | : | : | : | : |
 207 sgssvqwsdp1pecreiyc-----pappqldngliqgerdhgyrgsvtyacnkgftmig 261

QY 169 DEEMHC---SDGFWSEKPKC-----VEISCKSPDVIN-----GSPISQK 206
 Db : : | : | | | | | : | : | : | : | : | : | : | : | : | : |

RESULT 9
 AAP94773
 ID AAP94773 standard; protein; 381 AA.
 AC AAP94773;
 XX
 XX 04-JUL-1990 (first entry)
 XX
 DE Decay accelerating factor (DAF) of clones lambda 33 and lambda 47.
 XX
 XX DAF; allograft rejection; affinity purification;
 KW autoimmune disease; ds.
 XX
 OS Synthetic.
 XX
 PN W08901041-A.
 XX
 XX 09-FEB-1989.
 XX
 PF 03-AUG-1988; 88WO-US02648.
 XX
 PR 06-AUG-1987; 87US-0083757.
 XX
 XX (GETH) GENETECH INC.
 XX
 XX Caras I;
 XX
 XX WPI; 1989-061177/08.
 DR N-PSDB; AAN91043.
 XX
 PT Fusion polypeptide for targeting protein to cell membrane -

PT comprisesphospholipid anchor domain with heterologous
 PT polypeptide.
 XX
 PS Disclosure; ; 61pp; English.
 XX
 CC Recombinant DAF's are useful in treatment of inflammatory or cell lytic
 CC autoimmune diseases and allograft rejection. Useful in diagnostic
 CC compositions or in affinity purification.
 XX
 SQ Sequence 381 AA;

Query Match 22.4%; Score 261; DB 10; Length 381;
 Best Local Similarity 29.9%; Pred. No. 4.1e-16;
 Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

QY 1 EDCN---ELPPRRNTEILTGSHSDQTY-PEGTQAIYKCRPGYR---SLGNVIMVCRKGEW 53
 Db | | | | | : | : | | | | | | | | | | : | : |
 92 efncrscevpnlrnsaslkqpyitqnyfpvgtvveecrpgyrrepslsplktclqnlkw 151

QY 54 VALNPLRKCOKRCPGHPGDTPEFTLTGNGVFEYGVKAVYTCNEGYQLLGEINRECDT 113
 Db : | | : | | | : | : | | | : | : | | | | : | : |
 152 --stavefckkscnpngeirngqidvpqglf--gatisfscntgyklfgsts-sfc11 206

QY 114 DG----WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRFVCSGYKIEG 168
 Db | | | : | | : | : | : | : | : | : | : | : | : | : |
 207 sgssvqwsdp1pecreiyc-----pappqldngliqgerdhgyrgsvtyacnkgftmig 261

QY 169 DEEMHC---SDGFWSEKPKC-----VEISCKSPDVIN-----GSPISQK 206
 Db : : | : | | | | | : | : | : | : | : | : | : | : | : | : |

RESULT 10
 AAR66683
 ID AAR66683 standard; Protein; 381 AA.
 AC AAR66683;
 XX
 XX 23-JUL-1995 (first entry)
 XX
 DE Decay accelerating factor.
 XX
 KW Decay accelerating factor; DAF; mDAF; fusion protein; liposome;
 KW cell targeting; glycoposphatidylinositol; GPI; drug delivery.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..34
 FT /label= sig_peptide
 FT Modified-site 364
 FT /note= "probable phosphatidylinositol
 FT derivatization site".
 XX
 PN US5374548-A.
 XX
 XX 20-DEC-1994.
 XX
 PF 02-MAY-1986; 86US-0859107.
 XX
 PR 02-MAY-1986; 86US-0859107.
 PR 06-AUG-1987; 87US-0083757.
 PR 19-DEC-1991; 91US-0811048.
 PR 12-FEB-1993; 93US-0017934.
 XX
 XX (GETH) GENENTECH INC.
 XX
 XX Caras IW;
 XX
 XX WPI; 1995-035649/05.
 DR N-PSDB; AAR66683.
 XX

XX 26-1

PR 02-MAY-1986; 86US-0859107.

XX (GETH) GENENTECH INC.

XX Caras IW;

XX WPI; 1987-308481/44.

DR N-PSDB; AAN70048.

XX New decay accelerating factor variants - obtained with the factor by
PT using recombinant DNA procedures.

XX Disclosure; Page 18-20; 20pp; English.

XX The protein sequence is a variant of decay acceleration factor, sDAF.
CC DAF and variants are useful for treating paroxysmal nocturnal
CC haemoglobinuria, or inflammatory or cell lytic autoimmune diseases.
CC They may be used to ameliorate allograft rejection or autoimmune
CC diseases.

CC See also AAN70046 and AAN70047.

XX Sequence 440 AA;

Query Match 22.4%; Score 261; DB 8; Length 440;

Best Local Similarity 29.9%; Pred. No. 4.9e-16;

Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

QY 1 EDCN---ELPPRRNTEILTGSWSQTY-PEGTQAIYKCRPGYR---SLGNVIMVCRKGEW 53

Db 92 efcnrscveptrlnsaslkqpyitqnyfpvgtvveyecrpyrrepslspkltclqlnkw 151

QY 54 VALNPLRCQKRPCGHPGDTPTGTTLTGGNVFEYGVKAVYTCNMGYQLLGEINRECDT 113

Db 152 --stavefckkkscpnpgelrngqidvpqgilf--gatisfscntgylfgsts-sfcll 206

QY 114 DG---WTNDIPICEVVKCLPVTAPENKIVSSAMEPDRE-YHFGQAVRFVCSNGYKIEG 168

Db 207 sgssvqwsdplpecreiyc-----pappqldngliqgerdhgyrgsvtyacnkgftmig 261

QY 169 DEEMHC---SDDGFWSKEKPKC-----VEISCKSPDVIN-----GSPISQK 206

Db 262 ehslyctvnndegewsgpppecrgksltskvpptvqkpttnvpttevsptsqk 315

RESULT 15

AAP94774

ID AAP94774 standard; protein; 440 AA.

XX AAP94774;

XX 04-JUL-1990 (first entry)

XX Membrane bound decay accelerating factor (sDAF).

DE DAF; allograft rejection; affinity purification;
KW autoimmune disease; ds.

XX Homo sapiens.

XX WO8901041-A.

XX 09-FEB-1989.

XX 03-AUG-1988; 88WO-US02648.

XX 06-AUG-1987; 87US-0083757.

XX (GETH) GENENTECH INC.

XX Caras I;

XX WPI; 1989-061177/08.

DR N-PSDB; AAN91044.

XX Fusion polypeptide for targeting protein to cell membrane -
PT comprises phospholipid anchor domain with heterologous
PT polypeptide.

XX Disclosure; 2 61pp; English.

XX Recombinant DAF's are useful in treatment of inflammatory or cell lytic
CC autoimmune diseases and allograft rejection. Useful in diagnostic
CC compositions or in affinity purification.

XX Sequence 440 AA;

Query Match 22.4%; Score 261; DB 10; Length 440;

Best Local Similarity 29.9%; Pred. No. 4.9e-16;

Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;

QY 1 EDCN---ELPPRRNTEILTGSWSQTY-PEGTQAIYKCRPGYR---SLGNVIMVCRKGEW 53

Db 92 efcnrscveptrlnsaslkqpyitqnyfpvgtvveyecrpyrrepslspkltclqlnkw 151

QY 54 VALNPLRCQKRPCGHPGDTPTGTTLTGGNVFEYGVKAVYTCNMGYQLLGEINRECDT 113

Db 152 --stavefckkkscpnpgelrngqidvpqgilf--gatisfscntgylfgsts-sfcll 206

QY 114 DG---WTNDIPICEVVKCLPVTAPENKIVSSAMEPDRE-YHFGQAVRFVCSNGYKIEG 168

Db 207 sgssvqwsdplpecreiyc-----pappqldngliqgerdhgyrgsvtyacnkgftmig 261

QY 169 DEEMHC---SDDGFWSKEKPKC-----VEISCKSPDVIN-----GSPISQK 206

Db 262 ehslyctvnndegewsgpppecrgksltskvpptvqkpttnvpttevsptsqk 315

Search completed: August 29, 2002, 15:03:37

Job time: 168 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2002, 06:04:46 ; Search time 58.13 seconds
(without alignments)
438.047 Million cell updates/sec

Title: US-09-316-163-10

Perfect score: 1497

Sequence: 1 EDCELPPRRNTEILTGWS.....EKSCDNPYPNGDYSPLRIK 265

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71:**

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1497	100.0	449	1 NBHUHS	complement factor
2	1497	100.0	1231	1 NBHSH	complement factor
3	1024	68.4	1234	1 NBKSH	complement factor
4	699	46.7	669	2 S65551	factor H - bovine
5	440	29.4	1053	2 S46199	probable complement
6	332.5	22.2	360	2 T42921	complement control
7	303	20.2	597	1 NBHUC4	C4b-binding protein
8	302	20.2	597	1 S53711	C4BP alpha chain p
9	301.5	20.1	263	1 WMVZSP	apolipoprotein H h
10	301	20.1	676	2 A45900	complement C3b rec
11	298.5	19.9	2489	2 I73012	complement C3b/C4b
12	290.5	19.4	2014	2 I36936	complement recepto
13	289.5	19.3	497	2 JC2054	complement regulat
14	284.5	19.0	263	1 C36838	complement control
15	284.5	19.0	263	2 T28450	hypothetical prote
16	284.5	19.0	482	2 A34924	complement C3b/C4b
17	283.5	18.9	263	2 B72152	B18L protein - var
18	277.5	18.5	560	2 T16833	hypothetical prote
19	276.5	18.5	558	2 S57953	C4BP protein alpha
20	275	18.4	469	1 NBMSC4	C4b-binding protei
21	274.5	18.3	349	2 G02913	sperm CD46 - human
22	274.5	18.3	369	2 I57998	membrane cofactor
23	274	18.3	377	2 I54479	membrane cofactor
24	274	18.3	384	2 S01896	membrane cofactor
25	272	18.2	610	2 A35046	E-selectin precurs
26	270.5	18.1	369	2 JC5194	membrane cofactor
27	270.5	18.1	362	2 JC5138	membrane cofactor
28	270.5	18.1	440	2 A43519	complement recepto
29	266	17.8	482	2 JC5092	E-selectin - pig

30	265.5	17.7	610	1 I46001	C4b-binding protei
31	264.5	17.7	302	1 WMBE1E	secretory compleme
32	264.5	17.7	360	1 WMBE2E	membrane-bound com
33	262	17.5	381	1 B26359	decay-accelerating
34	262	17.5	440	2 A26359	decay-accelerating
35	260.5	17.4	768	2 A42755	P-selectin precurs
36	260	17.4	612	2 B42755	E-selectin precurs
37	257.5	17.2	1025	1 A43526	complement C3d/Eps
38	257	17.2	579	2 A56740	sperm-egg recognit
39	255	17.0	452	2 A35068	complement factor
40	254	17.0	768	2 I53821	P-selectin - rat
41	253	16.9	345	1 NBMS	apolipoprotein H p
42	252.5	16.9	1091	1 PL0009	complement C3d/Eps
43	252	16.8	340	2 I56234	decay-accelerating
44	252	16.8	661	1 KFHU13	coagulation factor
45	250.5	16.7	830	2 A30359	P-selectin precurs

ALIGNMENTS

RESULT 1

NBHUHS

complement factor H precursor, short splice form [validated] - human
N: Alternate names: complement factor H-related protein; complement protein H
C: Species: Homo sapiens (man)

C: Date: 31-Dec-1993 #sequence.revision 23-Feb-1996 #text.change 08-Dec-2000

C: Accession: S03013; B60238; A27877; A61103; A26505; S10479

R: Ripocher, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.

Biochem. J. 249, 593-602, 1988

A: Title: The complete amino acid sequence of human complement factor H.

A: Reference number: S00254; MUID: 88134059

A: Accession: S03013

A: Molecule type: mRNA

A: Residues: 1-449 <RIP>

A: Cross-references: EMBL:X07523; EMBL:Y00716; NID:g32492; PIDN:CAA30403.1; PID:g75807

A: Note: Part of this sequence, including the amino end of the mature protein was conf

A: Note: 402-Tyr was also found

R: Estaller, C.; Schwaebel, W.; Dierich, M.; Weiss, E.H.

Eur. J. Immunol. 21, 799-802, 1991

A: Title: Human complement factor H: two factor H proteins are derived from alternativ

A: Reference number: A60238; MUID: 91184292

A: Accession: B60238

A: Status: not compared with conceptual translation

A: Molecule type: mRNA

A: Residues: 1-33; 434-449 <EST>

A: Note: only portions of this 1.8 kilobase mRNA were sequenced

R: Schulz, T.F.; Schwaebel, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.

Eur. J. Immunol. 16, 1351-1355, 1986

A: Title: Human complement factor H: isolation of cDNA clones and partial cDNA sequenc

A: Reference number: A27877; MUID: 87054207

A: Accession: A27877

A: Molecule type: mRNA

A: Residues: 1-IL, 55-401, 'V', 403-449 <SCH>

A: Cross-references: GB:X04697; NID:g31991; PIDN:CAB41739.1; PID:g4725976

A: Note: an additional nucleotide present within the codon for Glu-310 was thought to

R: Schwaebel, W.; Zwirner, J.; Schulz, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.

Eur. J. Immunol. 17, 1485-1489, 1987

A: Title: Human complement factor H: expression of an additional truncated gene produc

A: Reference number: A61103; MUID: 88055295

A: Accession: A61103

A: Status: not compared with conceptual translation

A: Molecule type: mRNA

A: Residues: 27-76 <SC>

A: Note: this is a partial sequence of an alternatively spliced 1.8 kilobase mRNA that

R: Sim, R.B.; Discipio, R.G.

Biochem. J. 205, 285-293, 1982

A: Title: Purification and structural studies on the complement-system control protein

A: Reference number: A26505; MUID: 83048213

A: Accession: A26505

A: Molecule type: protein

A: Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>

R: Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P

Biochemistry 31, 3626-3634, 1992
A:Title: Solution structure of the fifth repeat of factor H: A second example of the complement factor H.
A:Reference number: A44551; MUID:92232649
A:Contents: annotation: NMR structure determination, residues 264-292
R:Kristensen, T.; Wetsel, R.A.; Tack, B.F.
J. Immunol. 136, 3407-3411, 1986
A:Title: Structural analysis of human complement protein H: homology with C4b binding protein.
A:Reference number: S10479; MUID:86169701
A:Accession: S10479
A:Molecule type: mRNA
A:Residues: 226-401, 'Y', 403-449 <KRI>
A:Cross-references: GB:M2383; NID:g180472; PID:AAAS2013.1; PID:g180473
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. Hc
C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.
C:Genetics: <HF1>
A:Gene: GDB:HF1; HF
A:Cross-references: GDB:120041; OMIM:134370
A:Map position: 1q32-1q32
C:Genetics: <HF2>
A:Gene: GDB:HF2; HF
A:Cross-references: GDB:129095
A:Map position: 1q32-1q32
A:Note: the correspondence between the two loci and the sequences indicated is unclear;
C:Function:
A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increased
he alternative complement pathway
A:Pathway: complement alternate pathway
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-449/Product: complement factor H, short splice form #status experimental <MAT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357
F:217/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 100.0%; Score 1497; DB 1; Length 449;
Best Local Similarity 100.0%; Pred. No. 7.9e-106;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	EDCNELPPRRNTEILTGSDQTYPEGTQAIYKRCPCGYRSLGNVIMVCRKEWVALNPLR	60
Db	19	EDCNELPPRRNTEILTGSDQTYPEGTQAIYKRCPCGYRSLGNVIMVCRKEWVALNPLR	78
QY	61	KCKRCPGHPGDPFGFTTGTGNVFEYGVKAVYTCNEGQYLLGEINRECDTDGWTNDI	120
Db	79	KCKRCPGHPGDPFGFTTGTGNVFEYGVKAVYTCNEGQYLLGEINRECDTDGWTNDI	138
QY	121	PICEVWKCLPVTAPENGKIVSSAMEPDREYHFGQAVRVCNSGYKIEGDEMHCSDDGF	180
Db	139	PICEVWKCLPVTAPENGKIVSSAMEPDREYHFGQAVRVCNSGYKIEGDEMHCSDDGF	198
QY	181	SKKPKCVCETSKSPDVINGSPISQKIYKENERFQYKCMNGYEYSGRDAVCTESGWRP	240
Db	199	SKKPKCVCETSKSPDVINGSPISQKIYKENERFQYKCMNGYEYSGRDAVCTESGWRP	258
QY	241	LPSCEEKSCDNPYPNGDYSPLRIK	265
Db	259	LPSCEEKSCDNPYPNGDYSPLRIK	283

RESULT 2

NBHUH

complement factor H precursor, long splice form [validated] - human
C:Species: Homo sapiens (man)

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Dec-2000

C:Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A:Title: The complete amino acid sequence of human complement factor H.
A:Reference number: S00254; MUID:88134059
A:Accession: S00254
A:Molecule type: mRNA
A:Residues: 1-1231 <RIP>
A:Cross-references: EMBL:Y00716; NID:g31964; PIDN:CAA68704.1; PID:g31965
A:Note: 402-Tyr was also found
A:Note: parts of this sequence, including the amino and carboxyl ends of the mature p
R:Estaller, C.; Schwaebble, W.; Dierich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A:Title: Human complement factor H: two factor H proteins are derived from alternativ
A:Reference number: A60238; MUID:91184292
A:Accession: A60238
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-56;1177-1231 <EST>
A:Note: only portions of this 4.3 kilobase mRNA were sequenced
R:Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.
Biochem. Rep. 7, 201-207, 1987
A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human compl
A:Reference number: A54726; MUID:88025472
A:Accession: A54726
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-579-1231 <DAY>
A:Cross-references: GB:M17517; NID:g180497; PIDN:AAAS2016.1; PID:g180498
A:Note: parts of this sequence were determined by protein sequencing
R:Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.
Biochem. Rep. 6, 65-72, 1986
A:Title: Partial characterization of human complement factor H by protein and cDNA se
A:Reference number: A61565; MUID:86188123
A:Accession: A61565
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1050-1057, 'T', 1059-1102 <RI2>
R:Sim, R.B.; DiScipio, R.G.
Biochem. J. 205, 285-293, 1982
A:Title: Purification and structural studies on the complement-system control protein
A:Reference number: A26505; MUID:83048213
A:Accession: A26505
A:Molecule type: protein
A:Residues: 19-20, 'Q', 22-23, 'V', 31-33, 'Q', 35 <SIM>
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, F
Biochemistry 31, 3626-3634, 1992
A:Title: Solution structure of the fifth repeat of factor H: A second example of the
A:Reference number: A44551; MUID:92232649
A:Contents: annotation: NMR structure determination, residues 264-292
R:Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.
J. Mol. Biol. 219, 717-725, 1991
A:Title: Three-dimensional structure of a complement control protein module in soluti
A:Reference number: A49224; MUID:91278097
A:Contents: annotation: NMR structure determination, residues 927-985
R:Estaller, C.; Kostiainen, V.; Schwaebble, W.; Dierich, M.P.; Weiss, E.H.
J. Immunol. 146, 3190-3196, 1991
A:Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a nc
A:Reference number: I56100; MUID:91201892
A:Accession: I72654
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1047-1231 <RES>
A:Cross-references: GB:M65294; NID:g183766; PIDN:AAA35948.1; PID:g183767
R:Carroll, J.A.; Bates, R.C.; Smith, A.I.; Tetoz, T.; Arellano, A.; Gordon, D.L.; Burn
Biochim. Biophys. Acta 1289, 305-311, 1996
A:Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.
A:Reference number: S66298; MUID:96205365
A:Accession: S66298
A:Status: preliminary
A:Molecule type: protein
A:Residues: 411-419,574-578,580-582 <CAR>

Query Match	100.0%;	Score 1497;	DB 1;	Length 1231;
Best Local Similarity	100.0%;	Pred. No. 2.3e-105;		
Matches 265;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	EDCNELPPRRNTEILTGSWSDDQTYPEGTQAIIYKCRPGYRSLGNVIMVCRKGWVALNPLR	60	
Dd				
	19	EDCNELPPRRNTEILTGSWSDDQTYPEGTQAIIYKCRPGYRSLGNVIMVCRKGWVALNPLR	78	
QY	61	KCOKRCGHPGDTPFGFTFLTGSNVEFYGVKAVYTCEGYQLLGEINYNRECDTDGWTNDI	120	
Dd				
	79	KCOKRCGHPGDTPFGFTFLTGSNVEFYGVKAVYTCEGYQLLGEINYNRECDTDGWTNDI	138	
QY	121	PICEVKLCPLVTAPENGGKIVSSAMEPDREWHFQGQAFVFCNSGYKIEGDEEMHCSDDGFW	180	
Dd				
	139	PICEVKLCPLVTAPENGGKIVSSAMEPDRETHFQGQAFVFCNSGYKIEGDEEMHCSDDGFW	198	
QY	181	SKEKPACVEISCKSPDIVNGSPISQKIIYKENERFOYKCNMGVEYSERGDAVCTESGWPR	240	
Dd				
	199	SKEKPACVEISCKSPDIVNGSPISQKIIYKENERFOYKCNMGVEYSERGDAVCTESGWPR	258	
QY	241	LPSCEEKSCDNPYPINGDYSPLRIK	265	
Dd				
	259	LPSCEEKSCDNPYPINGDYSPLRIK	283	

QY 1 EDCNELPRRNTETLTGSWSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEVALNPLR 60
 matches 111; Conservative 30; Mismatches 57; Indels 0; Gaps 0;

RESULT 3

NB!SH

complement factor H precursor - mouse

N:Alternate names: protein beta-1-H

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 22-Jun-1999

C:Accession: A26154; I49711; I49728

R:Kristensen, T.; Tack, B.F.

P:Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986

A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in length

A:Reference number: A26154; MUID:8623353

A:Accession: A26154

A:Molecule type: mRNA

A:Residues: 1-1234 <XRI>

A:Cross-references: GB:M12660; NID:g193724; PIDN:AAA37759.1; PID:g387181

R:Natsunome-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.

J. Immunol. 144, 358-362, 1990

A:Title: Demonstration of an unusual allelic variation of mouse factor H by the complement

A:Reference number: I49711; MUID:90111033

A:Accession: I49711

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-18 <RES>

A:Cross-references: GB:M31979; NID:g193726; PIDN:AAA37762.1; PID:g193729

R:Munoz-Canoves, P.; Tack, B.F.; Vik, D.P.

Biochemistry 28, 9891-9897, 1989

A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma

A:Reference number: I49728; MUID:90148935

A:Accession: I49728

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-19 <RES2>

A:Cross-references: GB:J02891; NID:g193805; PIDN:AAA37795.1; PID:g553926

C:Comment: Two codominant alleles of factor H are present in mice

C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine protease (C5 convertase) in the alternative complement pathway.

C:Genetics:

A:Map position: 1

C:Superfamily: complement factor H; complement factor H repeat homology

C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-1234/Product: complement factor H #status predicted <MPT>

F:21-80/Domain: complement factor H repeat homology <FH01>

F:85-141/Domain: complement factor H repeat homology <FH02>

F:146-205/Domain: complement factor H repeat homology <FH03>

F:210-262/Domain: complement factor H repeat homology <FH04>

F:246-248/Region: cell attachment (R-G-D) motif

F:267-320/Domain: complement factor H repeat homology <FH05>

F:325-385/Domain: complement factor H repeat homology <FH06>

F:389-442/Domain: complement factor H repeat homology <FH07>

F:448-505/Domain: complement factor H repeat homology <FH08>

F:509-564/Domain: complement factor H repeat homology <FH09>

F:569-622/Domain: complement factor H repeat homology <FH10>

F:629-683/Domain: complement factor H repeat homology <FH11>

F:690-743/Domain: complement factor H repeat homology <FH12>

F:752-802/Domain: complement factor H repeat homology <FH13>

F:808-861/Domain: complement factor H repeat homology <FH14>

F:867-931/Domain: complement factor H repeat homology <FH15>

F:936-989/Domain: complement factor H repeat homology <FH16>

F:1054-1048/Domain: complement factor H repeat homology <FH17>

F:1053-1107/Domain: complement factor H repeat homology <FH18>

F:1114-1168/Domain: complement factor H repeat homology <FH19>

F:1172-1233/Domain: complement factor H repeat homology <FH20>

F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,0-802,808-850,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,10

F:676,721,773,801,1030,1061,1225/Binding site: carbohydrate (Asn) (covalent) #status

Db 19 EDCKGPPPRENSSEILSGSWSEQLYPECTQATYKCRPGYRTLTGTVKCKNGKVVASPSR 78
QY 61 KCKRCRPGHGDTPFGTFTLTGNNVFEYGVKAVYTCNEGQVLLGEINRYRCDDTGDWTNDI 120
Db 79 ICRKKKCGHGDTPFGSFRVAVSQFEFGAKVYTCDDGQVLLGEIDYRCGADGWINDI 138
QY 121 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRVCNSGYKTEGDEMHCSDDGFW 180
Db 139 PICEVVKCLPVTLENGRIVSGRAETDQYFFGVQVRFECNSGFKTEGHEIHCSENGLW 198
QY 181 SKKPKCVETSCSDPDVINGSPIQIIYKENERFQYKCMNGYEYSRGGDAVCTESGWRP 240
Db 199 SNEKPRCVETLTPPRVENGDIINVPVYKENERYHYKCKHGYVPRKRGDAVCTGSGWS 258
QY 241 LPSCEEKSCDNPYPNGDYSPLR 264
Db 259 QPCEERKSPYILNGIYTPHRI 282

RESULT 4
S46199
factor H - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C>Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S65551
R:Soames, C.J.; Day, A.J.; Sim, R.B.
Biochem. J. 315, 523-531, 1996
A:Title: Prediction from sequence comparisons of residues of factor H involved in the in
A:Reference number: S65551; MUID:96202005
A:Accession: S65551
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-669 <SOA>
A:Cross-references: GB:X98697; NID:g1419423; PIDN:CAA67257.1; PID:g1419424
C:Superfamily: complement factor H; complement factor H repeat homology
F:55-114/Domain: complement factor H repeat homology <FHR1>
F:296-349/Domain: complement factor H repeat homology <FHR2>
F:355-412/Domain: complement factor H repeat homology <FHR3>
F:416-471/Domain: complement factor H repeat homology <FHR4>
F:476-530/Domain: complement factor H repeat homology <FHR5>
F:538-592/Domain: complement factor H repeat homology <FHR6>
F:599-651/Domain: complement factor H repeat homology <FHR7>

Query Match 46.7%; Score 699; DB 2; Length 669;
Best Local Similarity 63.4%; Pred. No. 2.6e-45;
Matches 118; Conservative 26; Mismatches 42; Indels 0; Gaps 0;

QY 76 GTFTLTGNNVFEYGVKAVYTCNEGQVLLGEINRYRCDDTGDWTNDIPICEVVKCLPVTAP 135
Db 3 GSPHLAEGNOFEYGAQVYTCDDGQVLLGEINRYRCDDTGDWTNDIPICEVVKCLPVTAP 62
QY 136 NGKIVSSAMEPDREYHFGQAVRVCNSGYKTEGDEMHCSDDGFWSEKPKCVETSCSKSP 195
Db 63 NGKIFSDALEPDQYFYGVQVFECSNGYMLDGPKHCSAGVWSEATPKCVETICKPP 122
QY 196 DVINGSPISOKIIYKENERFQYKCMNGYEYSRGGDAVCTESGWRPLPSCEEKSCDNPYP 255
Db 123 VILNGQAVLPKATYKQNERVQVRCAGAFYGGQRTVCTKSGWTPAPTCTEITCDPPRIP 182
QY 256 NGDYSP 261
Db 183 NGVIRP 188

RESULT 5
S46199
probable complement regulatory plasma protein SBI - barred sand bass
C:Species: Paralabrax nebulifer
C>Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 18-Jul-2001
C:Accession: S46199; S77894
R:Dahmen, A.; Kaidoh, T.; Zipfel, P.F.; Gigli, I.

Biochem. J. 301, 391-397, 1994
A:Title: Cloning and characterization of a cDNA representing a putative complement-re
A:Reference number: S46199; MUID:94318039
A:Accession: S46199
A:Molecule type: mRNA
A:Residues: 1-1053 <DAH1>
A:Cross-references: EMBL:L21703; NID:g639894; PIDN:AAA92556.1; PID:g639895
A:Experimental source: liver
A:Accession: S77894
A:Molecule type: protein
A:Residues: 526-532, 'X', 534-537; 809-817, 'X', 819-826 <DAH2>
C:Genetics:
A:Gene: SBI
C:Superfamily: complement factor H repeat homology
C:Keywords: glycoprotein
F:89-145/Domain: complement factor H repeat homology <FH01>
F:334-389/Domain: complement factor H repeat homology <FH02>
F:450-502/Domain: complement factor H repeat homology <FH03>
F:569-624/Domain: complement factor H repeat homology <FH04>
F:682-738/Domain: complement factor H repeat homology <FH05>
F:743-802/Domain: complement factor H repeat homology <FH06>
F:935-989/Domain: complement factor H repeat homology <FH07>
F:993-1052/Domain: complement factor H repeat homology <FH08>

Query Match 29.4%; Score 440; DB 2; Length 1053;
Best Local Similarity 36.2%; Pred. No. 1.6e-25;
Matches 89; Conservative 35; Mismatches 108; Indels 14; Gaps 6;

QY 21 DQYPEGTQAIYKCRPGYRSLGNVIMVCRKGWVALNPLRCKCRKPGHGDTPFGFTTL 80
Db 47 EASYPGGGRVRCVNGYS--GFFKLVCEVGKMETRG--AKCQPRSCGHPGDAQFADPHL 102
QY 81 TGCNVREYGVKAVYTCNEGQVLLGEINRYRCDDTGDWTNDIPICEVVKCLPVTAPENGKIV 140
Db 103 AEGNDVFGSKVYTCQKGYQMVSRINRYRCVAGMDGVVYVPCESQCCQLIHVDNNVQVI 162
QY 141 SSAMEPDREYHFGQAVRVCNSGYKI--EGDEEMHCSDDGFWSEKPKCVETSCSKSPDVIN 199
Db 163 GG----PEEATFGNVYRFSCKSRSEILDGSPELYCDERGDWSPVVKCAITCAIPPIEN 218
QY 200 GSPISOKIIYKENERFQYKCMNGYEYSRGGDAVCTESG----WRPLPSCEEKSCDNPYP 255
Db 219 GNVPGAIRYKENDVLHYECDRAFKHIDR--PSTCIKQIKAEWSPTPLCESIKRLTIMD 277
QY 256 NGDYSP 261
Db 278 GTRYEP 283

RESULT 6
T42921
complement control protein homolog ccph - ateline herpesvirus 3 (strain 73)
C:Species: ateline herpesvirus 3
A:Variety: strain 73
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42921
R:Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A:Description: Primary structure of the herpesvirus ateles genome.
A:Reference number: Z22274
A:Accession: T42921
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-360 <ALB>
A:Cross-references: EMBL:AF083424; PIDN:AC95530.1
A:Experimental source: strain 73
C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom

Query Match 22.2%; Score 332.5; DB 2; Length 360;
Best Local Similarity 34.9%; Pred. No. 6.9e-18;
Matches 96; Conservative 32; Mismatches 106; Indels 41; Gaps 18;

Db 119 --LRNQVEIKTDLSPGSGQIEFSCSGFELIGTSTSR-CEVDQDGVGWSHPLPQCEIVK 175
QY 129 LPVTAPENGKIVSSAMEPDREYHFQAVRVCSGVKIEGDEEMHCSD-...GFWSEK 184
Db 176 KPPDRLNRHSGE-...ENFAYGFSVYSCDPRESLILGHASICTVNETIGVWRPSP 231
QY 185 PKCVELSKSPDVINGSPISO-KIIYKENERFOYKCNMGYSEYERGDV-C-TESGWRP 240
Db 232 PTCEKITCRKPDVSHGEMVSGFGPIYNYKDTIVFKCKQKQFVL--RGSSVIHCADSKWNP 289
QY 241 L-PSCEEKSCDN-PYIPNGDY 259
Db 290 SPFACEPNSCINLPDIPHASW 310

RESULT 8
S53711
C4BP alpha chain precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S53711
R:de Frutos, P.G.; Dahlbaeck, B.
Biochim. Biophys. Acta 1261, 285-289, 1995
A:Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence mo
A:Reference number: S53711; MUID:95226458
A:Accession: S53711
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-597 <DEF>
A:Cross-references: EMBL:235490
F:50-107/Domain: complement factor H repeat homology <FH1>
F:112-169/Domain: complement factor H repeat homology <FH2>
F:174-234/Domain: complement factor H repeat homology <FH3>
F:239-294/Domain: complement factor H repeat homology <FH4>
F:299-360/Domain: complement factor H repeat homology <FH5>
F:364-422/Domain: complement factor H repeat homology <FH6>
F:426-480/Domain: complement factor H repeat homology <FH7>
F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match 20.2%; Score 302; DB 1; Length 597;
Best Local Similarity 32.0%; Pred. No. 2,4e-15;
Matches 87; Conservative 37; Mismatches 118; Indels 30; Gaps 16;

QY 2 DNELPPTNTTEILGWSWDQYPTGTOAIYKRCPCYSLG-NVIMVCR-KGEWALNPL 59
Db 49 DCGP-PPHLLFASSISELSENEYQTILKYTCRPGYTRNGLNPLILCKPRLW---SVD 104
QY 60 RKCKRCPGHPGDTPTGFTLTGGNVFEGYKAVYTCNEGQYLLGEINYNRECDTD-...G 115
Db 105 TFCVKKRCRNPGLDLPNGQVEK--TDFSGSQIEFSCSEGYILIGSTT-SHCDI0EKGV 161
QY 116 WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYH-FQAVRVCSGVKIEGDEEMHC 174
Db 162 WSDPLPKCEIVKCEPPNPLNGKHNG-...NEDIHTYSSVYSCNPRFSLIGEAISIC 217
QY 175 S-----DDGFWSEKPKCVELSKSPDVINGSPISO-KIIYKENERFOYKCNMGYSEYER 229
Db 218 TVKNKTGVWSPSPVPCKEILICSPNPVPHGKIISGFGPIYNYKDSIMVTCIDGFLV--RG 275
QY 230 DAV--C-TESGWRPLPS-CEEKSC-DNPYIPN 256
Db 276 SSLHCELDKSNFSPVPVCSNCLGLPLNVP 307

RESULT 9

WNVZSP

apolipoprotein H homolog precursor - vaccinia virus

N:Alternate names: 35k secretory protein; C3L protein; virokin

C:Species: vaccinia virus

A:Note: host Homo sapiens (man)

C>Date: 31-Dec-1989 #sequence_revision 30-Jun-1990 #text_change 22-Jun-1999

C:Accession: A31005; B42504
R:Kotwal, G.J.; Moss, B.
Nature 335, 176-178, 1988
A:Title: Vaccinia virus encodes a secretory polypeptide structurally related to compl
A:Reference number: A31005; MUID:88318974
A:Accession: A31005
A:Molecule type: DNA
A:Residues: 1-263 <KOT>
A:Cross-references: GB:X13166; NID:g60690; PIDN:CAA31564.1; PID:g60691
A:Experimental source: strain WR
R:Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 517-563, 1990
A:Title: Appendix to "The complete DNA sequence of vaccinia virus".
A:Reference number: A42501
A:Accession: B42504
A:Molecule type: DNA
A:Residues: 1-263 <GOE>
A:Cross-references: GB:M35027; NID:g335317; PIDN:AAA47997.1; PID:g335345
A:Experimental source: strain Copenhagen
R:Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 247-266, 1990
A:Title: The complete DNA sequence of vaccinia virus.
A:Reference number: A42531; MUID:91021027
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom
C:Keywords: duplication; extracellular protein
F:1-19/Domain: signal sequence status predicted <SIG>
F:20-263/Product: C4b-binding protein homolog #status predicted <SIG>
F:21-81/Domain: complement factor H repeat homology <FH1>
F:86-143/Domain: complement factor H repeat homology <FH2>
F:148-201/Domain: complement factor H repeat homology <FH3>
F:206-261/Domain: complement factor H repeat homology <FH4>

Query Match 20.1%; Score 301.5; DB 1; Length 263;
Best Local Similarity 33.3%; Pred. No. 1.1e-15;
Matches 78; Conservative 29; Mismatches 106; Indels 21; Gaps 12;

QY 20 SDQTYPEGTQAIYKCRPGYR--SLGNVIMVCRKGEWALNPLRKCKRCPGHPGDTPTG 77
Db 40 ANANTNIGDTIETLCLPGYKRMGMPIYAKCTGTGTLFN---QCIRKRCSPRDIDNG 96
QY 78 FTLTGNTVEYGVKAVYTCNEGQYLLGEI-NYRECDTDG---WTNDIPICEVVKCLPVT 133
Db 97 LDIGG--VDFGSSITYSCNSGYHLIGESKVCYELGSGTGMVNPAPICESVKCQSPS 153
QY 134 PENGKIVSAMPEPDREYHFQAVRVCSGVKIEGDEEMHCSDDGFWSEKPKCVELSK 193
Db 154 ISNGR--HNGYED--FYTDGSVVTYSCNSGYSLIGNSVLCS-GGEWS-DPPTCQIVKCP 207
QY 194 SPDVINGSPISO-KIIYKENERFOYKCNMGYSEYERGDVAVCTESG-WRP-LPSC 244
Db 208 HPTISNGYLSGFKRSYSYNVDNDFKRYGLSGSSSTCSPGNTWKPELPKC 261

RESULT 10

A45900

complement C3b receptor type 2 long form precursor - mouse (fragment)
C:Species: Mus musculus (house mouse)

C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999

C:Accession: A45900; I48306

R:Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.

J. Immunol. 144, 3581-3591, 1990

A:Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2
A:Reference number: A45900; MUID:90229754

A:Accession: A45900

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-676 <KUR>

A:Cross-references: GB:M36470

A:Experimental source: clone 31-1

R:Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers,

A:Residues: 1-683,'X',685-894,'A',896-1000,1451-1471,'X',1473-2489 <VIK2>
A:Cross-references: GB:L17418; NID:9306678; PIDN:AAB0694.1; PID:9306680
R:Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.;
J. Exp. Med. 169, 847-863, 1989
A:Title: Structure of the human Crl gene. Molecular basis of the structural and quant
A:Reference number: A47602; MUID:89176869
A:Accession: A47602
A:Molecule type: DNA
A:Residues: 1-41 <WON>
A:Cross-references: GB:X14893
R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.
J. Exp. Med. 168, 1255-1270, 1988
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b r
type 1.
A:Reference number: S03291; MUID:89010527
A:Accession: S03291
A:Molecule type: mRNA
A:Residues: 26-584 <HOU>
A:Cross-references: EMBL:X14362; NID:930197; PIDN:CAA32541.1; PID:g736240
R:Klickstein, L.B.; Bartow, T.J.; Milletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.
J. Exp. Med. 168, 1699-1717, 1988
A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4b r
A:Reference number: S03843; MUID:89035992
A:Accession: S03843
A:Molecule type: not shown
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-894,'A',896-1000,1451-2064,'I',2066-2276,'P',2278-2299,'H',2301-2325,'T',2328-2332
A:Cross-references: EMBL:Y00816; NID:930185; PIDN:CAA68755.1; PID:g30186
R:Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Weis, J.H.; Wilson, J.G.; Fearon, D.T.
J. Exp. Med. 165, 1095-1112, 1987
A:Title: Human C3b/C4b receptor (CRL). Demonstration of long homologous repeating dom
A:Reference number: A28507; MUID:87168191
A:Accession: A28507
A:Molecule type: mRNA
A:Residues: 953-1221,'FV',1224-2064,'I',2066-2276,'P',2278-2299,'H',2301-2325,'T',2332-2338
A:Cross-references: GB:X05309; NID:930136; PIDN:CAA28933.1; PID:g809019
R:Wong, W.W.; Klickstein, L.B.; Smith, J.A.; Weis, J.H.; Fearon, D.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985
A:Title: Identification of a partial cDNA clone for the human receptor for complement
A:Reference number: A94073; MUID:86067975
A:Accession: A24748
A:Molecule type: mRNA
A:Residues: 311-333;729-745;831-845 <WO2>
A:Cross-references: GB:M11569; NID:g180991; PIDN:AAA52297.1; PID:g180995; GB:M11617;
C:Genetics:
A:Gene: GDB:CRL; CD35
A:Cross-references: GDB:l119800; OMIM:l20620
A:Map position: lq32-lq32
A:Introns: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2; 594/2; 609/2;
1484/2; 1531/1; 1646/1; 1705/1; 1739/2; 1768/1; 1844/1; 1904/1; 1937/2; 1966/1; 2099/2;
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: duplication; glycoprotein; transmembrane protein
F:1-41/Domain: signal sequence #status predicted <SIG>
F:42-2489/Product: complement C3b/C4b receptor, membrane-bound form #status predicted
F:42-584/Product: complement C3b/C4b receptor, secreted form #status predicted <MATS>
F:43-99/Domain: complement factor H repeat homology <FH01>
F:104-161/Domain: complement factor H repeat homology <FH02>
F:166-232/Domain: complement factor H repeat homology <FH03>
F:238-293/Domain: complement factor H repeat homology <FH04>
F:297-353/Domain: complement factor H repeat homology <FH05>
F:358-416/Domain: complement factor H repeat homology <FH06>
F:421-487/Domain: complement factor H repeat homology <FH07>
F:493-549/Domain: complement factor H repeat homology <FH08>
F:554-611/Domain: complement factor H repeat homology <FH09>
F:616-682/Domain: complement factor H repeat homology <FH10>
F:688-743/Domain: complement factor H repeat homology <FH11>
F:747-803/Domain: complement factor H repeat homology <FH12>
F:808-866/Domain: complement factor H repeat homology <FH13>
F:871-937/Domain: complement factor H repeat homology <FH14>
F:943-999/Domain: complement factor H repeat homology <FH15>
F:1004-1061/Domain: complement factor H repeat homology <FH16>
F:1066-1132/Domain: complement factor H repeat homology <FH17>

F;1138-1193/Domain: complement factor H repeat homology <FH18>
F;1197-1253/Domain: complement factor H repeat homology <FH19>
F;1258-1316/Domain: complement factor H repeat homology <FH20>
F;1321-1387/Domain: complement factor H repeat homology <FH21>
F;1393-1449/Domain: complement factor H repeat homology <FH22>
F;1454-1511/Domain: complement factor H repeat homology <FH23>
F;1516-1582/Domain: complement factor H repeat homology <FH24>
F;1588-1643/Domain: complement factor H repeat homology <FH25>
F;1647-1703/Domain: complement factor H repeat homology <FH26>
F;1708-1766/Domain: complement factor H repeat homology <FH27>
F;1771-1837/Domain: complement factor H repeat homology <FH28>
F;1846-1902/Domain: complement factor H repeat homology <FH29>
F;1907-1964/Domain: complement factor H repeat homology <FH30>
F;1969-2035/Domain: complement factor H repeat homology <FH31>
F;2041-2096/Domain: complement factor H repeat homology <FH32>
F;2100-2156/Domain: complement factor H repeat homology <FH33>
F;2161-2219/Domain: complement factor H repeat homology <FH34>
F;2224-2290/Domain: complement factor H repeat homology <FH35>
F;2298-2354/Domain: complement factor H repeat homology <FH36>
F;2359-2415/Domain: complement factor H repeat homology <FH37>

Query Match 19.9%; Score 298.5; DB 2; Length 2489;
Best Local Similarity 27.0%; Pred. No. 2e-14;
Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;

QY 3 CNE---LPPRNTTEILGTSNSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPL 59
Db 43 CNAPEWLPFPARPTNLT-----DEFEFPGTYLNYECRPGYGRPFSSICLNKSNVMTGAKD- 97
QY 60 RKCKRCGCHPGDTPFGFTLTGNGVFEYGVKAVYTCNEGVLGGEINRYEC-----DTDG 115
Db 98 -RCRRKSCRNPDPVNGMVHVIK--IQFGSQIKYCTKGRLIGS-SSATCIIISGDTVI 153
QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYHFGQAVRVCVNSG-----YKIEG 168
Db 154 WDNETPICDRIPCGLPPTI-TNGDFISTNRE---NFHGSVVTYRCNPGSGRKFVFLVG 209
QY 169 DEEMHC--SDD--GFWSKERPKC-VEISCKSPDVINGSPIS-QKIIYKENERFOYKCNMG 222
Db 210 EPSIYCTSNDDQVGISGAPQCIIIPNKCTPPNVENGILVSDNLSLNFVEFRCPQG 269
QY 223 YEYSERGDVCTE-SGWRP-LPSCSE-----KSCDNP 252
Db 270 FVMKGPVRVKCOALNKWEPELPSCSRVCQPPDPVLAERTQDKDNFSPGQEVFYSCEP 329
QY 253 Y-----IPNGDYSP 261
Db 330 YDLRGAASMRCTPOGDWSP 348

RESULT 12
I36936
complement receptor 1 - chimpanzee (fragment)
C:Species: Pan troglodytes (chimpanzee)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: I36936; 136937
R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.
J. Immunol. 153, 691-700, 1994
A:Title: Primary sequence of an alternatively spliced form of CR1. Candidate for the 75,
A:Reference number: I36935; MUID:94292799
A:Accession: I36936
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2014 <BIR1>
A:Cross-references: GB:L24920; NID:g551564; PIDN:AAA51438.1; PID:g557725
A:Accession: I36937
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-397,1751-2014 <BIR2>
A:Cross-references: GB:L24921; NID:g557726; PIDN:AAA51439.1; PID:g557727
C:Genetics:
A:Gene: CR1

C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: duplication; glycoprotein; tandem repeat
F;18-74/Domain: complement factor H repeat homology <FH01>
F;79-136/Domain: complement factor H repeat homology <FH02>
F;141-207/Domain: complement factor H repeat homology <FH03>
F;213-268/Domain: complement factor H repeat homology <FH04>
F;272-328/Domain: complement factor H repeat homology <FH05>
F;333-391/Domain: complement factor H repeat homology <FH06>
F;396-462/Domain: complement factor H repeat homology <FH07>
F;470-526/Domain: complement factor H repeat homology <FH08>
F;539-586/Domain: complement factor H repeat homology <FH09>
F;722-778/Domain: complement factor H repeat homology <FH10>
F;1041-1107/Domain: complement factor H repeat homology <FH11>
F;1172-1228/Domain: complement factor H repeat homology <FH12>
F;1233-1291/Domain: complement factor H repeat homology <FH13>
F;1296-1362/Domain: complement factor H repeat homology <FH14>
F;1432-1489/Domain: complement factor H repeat homology <FH15>
F;1625-1681/Domain: complement factor H repeat homology <FH16>
F;1749-1815/Domain: complement factor H repeat homology <FH17>
F;1823-1879/Domain: complement factor H repeat homology <FH18>
F;1884-1940/Domain: complement factor H repeat homology <FH19>

Query Match 19.4%; Score 290.5; DB 2; Length 2014;
Best Local Similarity 26.6%; Pred. No. 6.5e-14;
Matches 85; Conservative 49; Mismatches 112; Indels 73; Gaps 18;

QY 3 CNE---LPPRNTTEILGTSNSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPL 59
Db 18 CNAPEWLPFPARPTNLT-----DEFEFPGTYLNYECRPGYGRPFSSICLNKSNVMTGAKD- 72
QY 60 RKCKRCGCHPGDTPFGFTLTGNGVFEYGVKAVYTCNEGVLGGEINRYEC-----DTDG 115
Db 73 -RCRRKSCRNPDPVNGMVHVI--KDQFGSQIKYCTKGRLIGS-SSATCIIISGDTVI 128
QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYHFGQAVRVCVNSG-----YKIEG 168
Db 129 WDNETPICDRIPCGLPPTI-TNGDFISTNRE---NFHGSVVTYRCNPGSGRKFVFLVG 184
QY 169 DEEMHC--SDD--GFWSKERPKC-VEISCKSPDVINGSPIS-QKIIYKENERFOYKCNMG 222
Db 185 EPSIYCTSNDDQVGISGAPQCIIIPNKCTPPNVENGILVSDNLSLNFVEFRCPQG 244
QY 223 YEYSERGDVCTE-SGWRP-LPSCSE-----KSCDNP 252
Db 245 FVMKGPVRVKCOALNKWEPELPSCSRVCQPPDPVLAERTQDKDNFSPGQEVFYSCEP 304
QY 253 Y-----IPNGDYSP 261
Db 305 YDLRGAASLRCTPOGDWSP 323

RESULT 13
JC2054
complement regulatory protein, 512 antigen precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 21-Jul-2000
C:Accession: JC2054; PC2027
R:Sakurada, C.; Seno, H.; Dohi, N.; Takizawa, H.; Nonaka, M.; Okada, H.
Biochem. Biophys. Res. Commun. 190, 819-826, 1994
A:Title: Molecular cloning of the rat complement regulatory protein, 512 antigen.
A:Reference number: JC2054; MUID:94161746
A:Accession: JC2054
A:Molecule type: mRNA
A:Residues: 1-497 <SAK>
A:Cross-references: GB:D42114; NID:g577641; PIDN:BAA07698.1; PID:g987688
A:Accession: PC2027
A:Molecule type: protein
A:Residues: 39-51 <SA2>
C:Comment: This protein plays a critical role in protection against complement mediated
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: glycoprotein; transmembrane protein
F;1-36/Domain: signal sequence #status predicted <SIG>

Qy 194 SPDVINGSPISQ-KIIYKENERFOYKCNMGYEYSERGDVAICTESG-WRP-LPSC 244
|::||| | | : | ||: | : | |
Db 208 HPTILNGYLSSGFKRYSYNNDVDFTCYGYKLGGSSSTCSPGNWTQPELPC 261

RESULT 15
T28450
hypothetical protein D15L - variola major virus
C:Species: variola major virus
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
R:Accession: T28450
R:Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Au Nature 366, 748-751, 1993
A>Title: Potential virulence determinants in terminal regions of variola smallpox vir A:Reference number: Z20488; MUID:94088747
A:Accession: T28450
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-263 <MAS>
A:Cross-references: EMBL:L22579; NID:g6233595; PIDN:AAA60760.1; PID:g438930
A:Experimental source: strain "Bangladesh-1975"
C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom

Query Match 19.0%; Score 284.5; DB 2; Length 263;
Best Local Similarity 32.1%; Pred. No. 2.1e-14;
Matches 75; Conservative 29; Mismatches 109; Indels 21; Gaps 11;

Qy 20 SDQTYPEQTATYKCRPGYR--SLGNVINVCRCGENVALPLRKCKQRCPCHGDTPEGT 77
::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 40 ANANYINGDTIEYLCPLPYGRKKOMGPYAKCTGTGWLTFN---OCIARRCPSPRXDINGH 96
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 78 FTLTGNNVFEXGVKAVYTCEGYQLLGEI-NYRECDTG---WTNDIPICEVVRCLPVTA 133
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 LDIGG---VDFGSITYSNSGYLLIGEYKSYCKLGSTGSMWNPKAPICESVKQLPPS 153
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 134 PENGKIVSAMEPDREYHFQAVRFVNCNSGYKIEGDEMHCSDDGFWSKPKVEISCK 193
||: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 154 ISNGRHNGY----NDFYTDGSYVTYSCNSGYSLIGNSGVLCS-GGEWS-NPPTCQIVKCP 207
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 194 SPDVINGSPISQ-KIIYKENERFOYKCNMGYEYSERGDVAICTESG-WRP-LPSC 244
|::||| | | : | ||: | : | |
Db 208 HPTILNGYLSSGFKRYSYNNDVDFTCYGYKLGGSSSTCSPGNWTQPELPC 261

Search completed: August 30, 2002, 06:18:08
Job time: 802 sec

GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2002, 06:17:06 ; Search time 28.81 Seconds
 (without alignments)
 356.150 Million cell updates/sec

Title: US-09-316-163-10
 Perfect score: 1497
 Sequence: 1 EDCHLPFRRTTEITGSWS.....EKSCDNPYPNGDYSPLRIK 265

Scoring table:
 BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1497	100.0	1231	1 CFAH_HUMAN	P08603 homo sapien
2	1024	68.4	1234	1 CFAH_MOUSE	P06909 mus musculus
3	723.5	48.3	685	1 CFAH_BOVIN	Q28085 bos taurus
4	303	20.2	597	1 C4BP_HUMAN	P04003 homo sapien
5	301.5	20.1	263	1 VCP_VACCV	P10998 vaccinia vi
6	298.5	19.9	2039	1 CRI_HUMAN	P17927 homo sapien
7	276.5	18.5	558	1 C4BP_RAT	Q63514 rattus norv
8	275	18.4	469	1 C4BP_MOUSE	P08607 mus musculus
9	274	18.3	377	1 MCP_HUMAN	P15329 homo sapien
10	272	18.2	610	1 LEM2_HUMAN	P16581 homo sapien
11	266	17.8	484	1 LEM2_PIG	P98110 sus scrofa
12	265.5	17.7	610	1 C4BP_BOVIN	Q01016 herpesvirus
13	264.5	17.7	360	1 CCPH_HSVSA	P08174 homo sapien
14	262	17.5	381	1 DAF_HUMAN	Q01102 mus musculus
15	260.5	17.4	768	1 LEM3_MOUSE	Q00690 mus musculus
16	260	17.4	612	1 LEM2_MOUSE	P19070 mus musculus
17	257.5	17.2	1025	1 CR2_MOUSE	P98106 rattus norv
18	254	17.0	768	1 LEM3_RAT	P20023 homo sapien
19	254	17.0	1033	1 CR2_HUMAN	Q01339 mus musculus
20	253	16.9	345	1 APOH_MOUSE	P49457 pongo pygma
21	252	16.8	340	1 DAF_PONPY	P05160 homo sapien
22	252	16.8	661	1 F13B_HUMAN	P16109 homo sapien
23	250.5	16.7	830	1 LEM3_HUMAN	P98109 ovis aries
24	249.5	16.7	769	1 LEM3_SHEEP	Q60401 cavia porce
25	246.5	16.5	507	1 DAF_CAVPO	P98105 rattus norv
26	241	16.1	549	1 LEM2_RAT	P33730 canis faml
27	241	16.1	611	1 LEM2_CANFA	P27113 oryctolagus
28	240	16.0	345	1 APOH_BOVIN	Q07968 mus musculus
29	239.5	16.0	551	1 LEM2_RABIT	Q09101 drosophila
30	234	15.6	668	1 F13B_MOUSE	Q61475 mus musculus
31	233.5	15.6	958	1 HIG_DROME	P02749 homo sapien
32	231.5	15.5	390	1 DAF1_MOUSE	
33	231	15.4	345	1 APOH_HUMAN	

ALIGNMENTS

```

RESULT 1
CFAH_HUMAN
ID CFAH_HUMAN STANDARD; PRT; 1231 AA.
AC P08603; Q14570; P78435; Q9NU86;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement factor H precursor (H factor 1).
GN HFI OR HF OR CFH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT Y-402.
RC TISSUE=Liver;
RX MEDLINE=88134059; PubMed=2963625;
RA Ripoché J., Day A.J., Harris T.J.R., Sim R.B.;
RT "The complete amino acid sequence of human complement factor H.";
RL Biochem. J. 249:593-602(1988).
RN [2]
RP SEQUENCE OF 53-445 FROM N.A.
RX MEDLINE=87054207; PubMed=2946589;
RA Schulz T.F., Schwaebble W., Stanley K.K., Weiss E., Dierich M.P.;
RT "Human complement factor H: isolation of cDNA clones and partial cDNA
sequence of the 38-kDa tryptic fragment containing the binding site
for C3b.";
RL Eur. J. Immunol. 16:1351-1355(1986).
RN [3]
RP SEQUENCE OF 226-449 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=86169701; PubMed=2937845;
RA Kristensen T., Wetsel R.A., Tack B.F.;
RT "Structural analysis of human complement protein H: homology with C4b
binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";
J. Immunol. 136:3407-3411(1986).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA Bird C.;
RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1047-1231 FROM N.A.
RX MEDLINE=91301892; PubMed=1826708;
RA Estaller C., Koistinen V., Schwaebble W., Dierich M.P., Weiss E.H.;
RT "Cloning of the 1.4-kb mRNA species of human complement factor H
reveals a novel member of the short consensus repeat family related
to the carboxy terminal of the classical 150-kDa molecule.";
J. Immunol. 146:3190-3196(1991).
RN [6]
RP SEQUENCE OF 19-35.
RX MEDLINE=83048213; PubMed=6215918;
RA Sim R.B., Discipio R.G.;
RT "Purification and structural studies on the complement-system control
protein beta 1H (Factor H).";
Biochem. J. 205:285-293(1982).
RN [7]
RP SEQUENCE OF 1-19 FROM N.A.

```

```

34 222.5 14.9 407 1 DAF2_MOUSE Q61476 mus musculus
35 222.5 14.9 485 1 LEM2_BOVIN P98107 bos taurus
36 222 14.8 345 1 APOH_CANFA P33703 canis faml
37 216.5 14.5 646 1 LEM3_BOVIN P42201 bos taurus
38 210 14.0 372 1 LEM1_MOUSE P18337 mus musculus
39 203.5 13.6 372 1 LEM1_HUMAN P14151 homo sapien
40 203.5 13.6 372 1 LEM1_PANTR Q95237 pan troglod
41 202.5 13.5 372 1 LEM1_PONPY P95235 pongo pygma
42 200 13.4 372 1 LEM1_RAT P30836 rattus norv
43 194 13.0 372 1 LEM1_MACMU Q95198 macaca mula
44 194 13.0 372 1 LEM1_PAPHA Q28768 papio hamad
45 192.5 12.9 270 1 FHR2_HUMAN P36980 homo sapien

```



```
Query Match 100.0%; Score 1497; DB 1; Length 1231;
Best Local Similarity 100.0%; Pred. No. 6.5e-117;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDNCNELPPRNTTELTGSDQTYPECTQAIYKRCGYRSLGNVIMVCRKGWVALNPLR 60
Db 19 EDNCNELPPRNTTELTGSDQTYPECTQAIYKRCGYRSLGNVIMVCRKGWVALNPLR 78
QY 61 KCQKRCGHPGDPFGFTLTGNGVEYGVKAVYTCNEGYYOLLGEINRYRCDDTGDGWTNDI 120
Db 79 KCQKRCGHPGDPFGFTLTGNGVEYGVKAVYTCNEGYYOLLGEINRYRCDDTGDGWTNDI 138
QY 121 PICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRFVNCNGYKTEGDEMHCSDDGFW 180
Db 139 PICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRFVNCNGYKTEGDEMHCSDDGFW 198
QY 181 SKKPKRVEISCKSPDVINGSPISQKIYKENERFOYKCNMGYEYSRGGDAVCTESGWRP 240
Db 199 SKKPKRVEISCKSPDVINGSPISQKIYKENERFOYKCNMGYEYSRGGDAVCTESGWRP 258
QY 241 LPSCEEKSCDNPYPNGDYSPLRIK 265
Db 259 LPSCEEKSCDNPYPNGDYSPLRIK 283

RESULT 2
CFAH_MOUSE STANDARD; PRT: 1234 AA.
AC P06909;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Complement factor H precursor (protein beta-1-H).
GN HF1 OR CFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86233353; PubMed=2940596;
RA Kristensen T., Tack B.F.;
RT "Murine protein H is comprised of 20 repeating units, 61 amino acids
in length.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).
RN [2]
RP SEQUENCE OF 1-19 FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=90148935; PubMed=2533512;
RA Munoz-Ganoves P., Tack B.F., Vik D.P.;
RT "Analysis of complement factor H mRNA expression: dexamethasone and
IFN-gamma increase the level of H in L cells.";
RL Biochemistry 28:9891-9897(1989).
RN [3]
RP SEQUENCE OF 1-18 FROM N.A.
RX MEDLINE=90111033; PubMed=2136885;
RA Natsume-Sakai S., Nonaka M., Nonaka M., Harada Y.N., Shreffler D.C.,
RA Moriwaki K.;
RT "demonstration of an unusual allelic variation of mouse factor H by
the complete cDNA sequence of the H.2 allotype.";
RL J. Immunol. 144:358-362(1990).
CC 1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3)NBB COMPLEX (C5
CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC 1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN
MICE.
CC 1- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M12660; AAA37759.1; -
CC EMBL: J02891; AAA37795.1; -
CC EMBL: M31979; AAA37762.1; -
CC PIR: A26154; NBM5H.
CC DR HSP; P08603; 1HF1.
CC DR MGD; MGI:88385; Cfh.
CC DR InterPro: IPR000436; Sushi_SCR_CCP.
CC Pfam: P00084; sushi; 20.
CC SMART: SM00032; CCP; 20.
CC Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW Signal.
FT SIGNAL 1 18 BY SIMILARITY.
FT CHAIN 19 1234 COMPLEMENT FACTOR H.
FT DOMAIN 20 81 SUSHI 1.
FT DOMAIN 84 142 SUSHI 2.
FT DOMAIN 145 206 SUSHI 3.
FT DOMAIN 209 263 SUSHI 4.
FT DOMAIN 266 321 SUSHI 5.
FT DOMAIN 324 386 SUSHI 6.
FT DOMAIN 388 443 SUSHI 7.
FT DOMAIN 447 506 SUSHI 8.
FT DOMAIN 508 565 SUSHI 9.
FT DOMAIN 568 623 SUSHI 10.
FT DOMAIN 628 684 SUSHI 11.
FT DOMAIN 689 744 SUSHI 12.
FT DOMAIN 751 803 SUSHI 13.
FT DOMAIN 807 862 SUSHI 14.
FT DOMAIN 866 932 SUSHI 15.
FT DOMAIN 935 990 SUSHI 16.
FT DOMAIN 993 1049 SUSHI 17.
FT DOMAIN 1052 1108 SUSHI 18.
FT DOMAIN 1113 1169 SUSHI 19.
FT DOMAIN 1171 1234 SUSHI 20.
FT DISULFID 21 66 BY SIMILARITY.
FT DISULFID 52 80 BY SIMILARITY.
FT DISULFID 85 129 BY SIMILARITY.
FT DISULFID 114 141 BY SIMILARITY.
FT DISULFID 146 192 BY SIMILARITY.
FT DISULFID 178 205 BY SIMILARITY.
FT DISULFID 210 251 BY SIMILARITY.
FT DISULFID 237 262 BY SIMILARITY.
FT DISULFID 267 309 BY SIMILARITY.
FT DISULFID 294 320 BY SIMILARITY.
FT DISULFID 325 374 BY SIMILARITY.
FT DISULFID 357 385 BY SIMILARITY.
FT DISULFID 389 431 BY SIMILARITY.
FT DISULFID 416 442 BY SIMILARITY.
FT DISULFID 448 494 BY SIMILARITY.
FT DISULFID 477 505 BY SIMILARITY.
FT DISULFID 509 553 BY SIMILARITY.
FT DISULFID 536 564 BY SIMILARITY.
FT DISULFID 569 610 BY SIMILARITY.
FT DISULFID 597 622 BY SIMILARITY.
FT DISULFID 629 672 BY SIMILARITY.
FT DISULFID 658 683 BY SIMILARITY.
FT DISULFID 690 732 BY SIMILARITY.
FT DISULFID 718 743 BY SIMILARITY.
FT DISULFID 752 791 BY SIMILARITY.
FT DISULFID 780 802 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 861 BY SIMILARITY.
FT DISULFID 867 920 BY SIMILARITY.
FT DISULFID 906 931 BY SIMILARITY.
FT DISULFID 936 978 BY SIMILARITY.
FT DISULFID 964 989 BY SIMILARITY.
FT DISULFID 994 1037 BY SIMILARITY.
FT DISULFID 1023 1048 BY SIMILARITY.
FT DISULFID 1033 1096 BY SIMILARITY.
```

```
FT DISULFID 1082 1107 BY SIMILARITY.
FT DISULFID 1114 1157 BY SIMILARITY.
FT DISULFID 1143 1168 BY SIMILARITY.
FT DISULFID 1172 1223 BY SIMILARITY.
FT DISULFID 1206 1233 BY SIMILARITY.
FT CARBOHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 773 773 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1225 1225 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1234 AA; 139082 MW; C5AC02F341B957F7 CRC64;

Query Match 68.4%; Score 1024; DB 1; Length 1234;
Best Local Similarity 67.0%; Pred. No. 1.5e-77;
Matches 177; Conservative 30; Mismatches 57; Indels 0; Gaps 0;

QY 1 EDCNELPPRRNTEILTGSDQTYPEGTQAIYKCRGYRSLGNVIMVCRKGEWVALNPLR 60
Db 19 EDCKGPPPRENSILSGNSEQLYPEGTQATYKCRGYRTLGTIVKCKNGKVASNP 78

QY 61 KCQKRCGHPGDPFGFTLTGCVNVEYGVKAVYTCNEGQVLLGEINRYRCDTGTNDI 120
Db 79 ICRKPCGHPGDPFGSFLAVGSQFEFGAKVYTCDDGYQLLGEIDYRCEGADGWINDI 138

QY 121 PICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRVCNSGYKIEGDEMHCSDDGF 180
Db 139 PICEVVKCLPVTLENGRIVSGAETDQYFGQVVRFCNSGFKIEGHEIHCSENGLW 198

QY 181 SKEPKKVEISCKSPDVINGSPIQKIYKENERFQYKCMNGYEYSGRDAVCTESGWRP 240
Db 199 SNEKPRCVELCTPPRVENGSGINVRKPVYKENERHYKCKHGYPVVRGSDAVCTGSWSS 258

QY 241 LPSCEEKSCDNPIPNQDYSPLRI 264
Db 259 QPFCEKRCSPPIVILNGIYTPHRI 282

RESULT 3
ID CFAH_BOVIN STANDARD; PRT; 685 AA.
AC Q28085;
OC 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Complement factor H (H factor 1) (Fragments).
GN HFI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 17-685 FROM N.A., AND SEQUENCE OF 1-16.
RC TISSUE=Liver;
RX MEDLINE=96202005; PubMed=8615824;
RA Soames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
in the interaction with complement component C3b.";
RL Blochem. J. 315:523-531(1996).
CC -I- FUNCTION: Factor H functions as a cofactor in the inactivation of
C3b by factor I and also increases the rate of dissociation of the
C3bb complex (C3 convertase) and the (C3b)NBB complex (C5
convertase) in the alternative complement pathway (By similarity).
CC -I- SIMILARITY: CONTAINS AT LEAST 13 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X98697; CAA67257.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 11.
DR SMART; SM00032; CCP; 11.
KW Complement alternate pathway; Plasma; Repeat; Sushi.
FT NON_TER 1
FT NON_CONS 16 17
FT DOMAIN <17 67 SUSHI 2.
FT DOMAIN 70 131 SUSHI 3.
FT DOMAIN 134 188 SUSHI 4.
FT DOMAIN 191 246 SUSHI 5.
FT DOMAIN 249 309 SUSHI 6.
FT DOMAIN 311 366 SUSHI 7.
FT DOMAIN 370 429 SUSHI 8.
FT DOMAIN 431 488 SUSHI 9.
FT DOMAIN 491 547 SUSHI 10.
FT DOMAIN 553 609 SUSHI 11.
FT DOMAIN 614 668 SUSHI 12.
FT DOMAIN 675 >685 SUSHI 13.
FT DISULFID 39 66 BY SIMILARITY.
FT DISULFID 71 117 BY SIMILARITY.
FT DISULFID 103 130 BY SIMILARITY.
FT DISULFID 135 176 BY SIMILARITY.
FT DISULFID 162 187 BY SIMILARITY.
FT DISULFID 192 234 BY SIMILARITY.
FT DISULFID 219 245 BY SIMILARITY.
FT DISULFID 250 297 BY SIMILARITY.
FT DISULFID 302 308 BY SIMILARITY.
FT DISULFID 312 354 BY SIMILARITY.
FT DISULFID 339 365 BY SIMILARITY.
FT DISULFID 371 417 BY SIMILARITY.
FT DISULFID 400 428 BY SIMILARITY.
FT DISULFID 432 476 BY SIMILARITY.
FT DISULFID 459 487 BY SIMILARITY.
FT DISULFID 492 534 BY SIMILARITY.
FT DISULFID 520 546 BY SIMILARITY.
FT DISULFID 554 597 BY SIMILARITY.
FT DISULFID 583 608 BY SIMILARITY.
FT DISULFID 615 656 BY SIMILARITY.
FT DISULFID 642 667 BY SIMILARITY.
FT NON_TER 685
SQ SEQUENCE 685 AA; 77536 MW; 69FC9DC8D530E872 CRC64;

Query Match 48.3%; Score 723.5; DB 1; Length 685;
Best Local Similarity 49.8%; Pred. No. 7.7e-53;
Matches 130; Conservative 28; Mismatches 46; Indels 57; Gaps 1;

QY 1 EDCNELPPRRNTEILTGSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLR 60
Db 1 EDCKEPPPRKTEILSVS ----- 18

QY 61 KCQKRCGHPGDPFGFTLTGCVNVEYGVKAVYTCNEGQVLLGEINRYRCDTGTNDI 120
Db 19 -----GSPHLAEGNQFEYGAQVYTCDEGYQMGEMNFRECDTGTNDI 63

QY 121 PICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRVCNSGYKIEGDEMHCSDDGF 180
Db 64 PICEVVKCLPVTPEPKGKIFSDALEPDQYTYGVQVQFECNSGYMLDGPQIHCSAGVW 123

QY 181 SKEPKKVEISCKSPDVINGSPIQKIYKENERFQYKCMNGYEYSGRDAVCTESGWRP 240
Db 124 SAETPRKVEIFCKPPVILNGQAVLPKATYKQNERVQYRCAAGFEYQGRGDTVCTKSGWTP 183

QY 241 LPSCEEKSCDNPIPNQDYSPLRI 261
Db 184 APTCIEITCDPPRIPNGVYRP 204
```


RESULT 4
C4BP_HUMAN
ID C4BP_HUMAN STANDARD: PRT; 537 AA.
AC P04003;
DT 23-OCT-1986 (Rel. 02, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein alpha chain precursor (C4bp) (Proline-rich
DE protein) (PRP).
GN C4BPA OR C4BP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=90073699; PubMed=2590215;
RA Matsuguchi T., Okamura S., Aso T., Sata T., Niho Y.;
RT "Molecular cloning of the cDNA coding for proline-rich protein (PRP):
RL Identity of PRP as C4b-binding protein.";
RL Biochem. Biophys. Res. Commun. 165:138-144(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91113199; PubMed=1989602;
RA Aso T., Okamura S., Matsuguchi T., Sakamoto N., Sata T., Niho Y.;
RT "Genomic organization of the alpha chain of the human C4b-binding
RL protein gene.";
RL Biochem. Biophys. Res. Commun. 174:222-227(1991).
RN [3]
RP SEQUENCE OF 9-81 FROM N.A.
RX MEDLINE=98242821; PubMed=3378624;
RA Lintin S.J., Lewin A.R., Reid K.B.M.;
RT "Derivation of the sequence of the signal peptide in human
RL C4b-binding protein and interspecies cross-hybridisation of the C4bp
RL cDNA sequence.";
RL FEBS Lett. 232:328-332(1988).
RN [4]
RP SEQUENCE OF 203-288 FROM N.A.
RX MEDLINE=86301119; PubMed=3017751;
RA Lintin S.J., Reid K.B.M.;
RT "Studies on the structure of the human C4b-binding protein gene.";
RL FEBS Lett. 204:77-81(1986).
RN [5]
RP SEQUENCE OF 80-597 FROM N.A.
RX MEDLINE=86025405; PubMed=3840370;
RA Chung L.P., Bentley D.R., Reid K.B.M.;
RT "Molecular cloning and characterization of the cDNA coding for C4b-
RL binding protein, a regulatory protein of the classical pathway of the
RL human complement system.";
RL Biochem. J. 230:133-141(1985).
RN [6]
RP SEQUENCE OF 49-88.
RX MEDLINE=85296001; PubMed=4033666;
RA Chung L.P., Gagnon J., Reid K.B.M.;
RT "Amino acid sequence studies of human C4b-binding protein: N-terminal
RL sequence analysis and alignment of the fragments produced by limited
RL proteolysis with chymotrypsin and the peptides produced by cyanogen
RL bromide treatment.";
RL Mol. Immunol. 22:427-435(1985).
RN [7]
RP ELECTRON MICROSCOPY, 3D-STRUCTURE, AND LIGAND BINDING.
RX MEDLINE=83221615; PubMed=6222381;
RA Dahlback B., Smith C.A., Mueller-Eberhard H.J.;
RT "Visualization of human C4b-binding protein and its complexes with
RL vitamin K-dependent protein S and complement protein C4b.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S

CC AND WITH SERUM AMYLOID P COMPONENT.
CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC OF 3 POSSIBLE SORTS: A 570 kDa COMPLEX OF 7 ALPHA CHAINS AND 1
CC BETA CHAIN, A 530 kDa HOMOPOLYMER OF ALPHA CHAINS OR A 500 kDa
CC COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
CC THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
CC BINDING SITE FOR C4B AT THE END.
CC -1- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31452; AAA36507.1; -;
CC EMBL; M62486; AAA36506.1; -;
CC EMBL; M62475; AAA36506.1; JOINED.
CC EMBL; M62476; AAA36506.1; JOINED.
CC EMBL; M62477; AAA36506.1; JOINED.
CC EMBL; M62478; AAA36506.1; JOINED.
CC EMBL; M62479; AAA36506.1; JOINED.
CC EMBL; M62480; AAA36506.1; JOINED.
CC EMBL; M62481; AAA36506.1; JOINED.
CC EMBL; M62482; AAA36506.1; JOINED.
CC EMBL; M62484; AAA36506.1; JOINED.
CC EMBL; X07853; CAA30701.1; -;
CC EMBL; X04284; CAB51244.1; -;
CC EMBL; X04296; CAA27839.1; -;
CC EMBL; X02865; CAA26617.1; -;
CC PIR; A33568; NBHUC4.
CC HSP; P10998; LVUD.
CC MIM; 120830; -;
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00084; sushi; 8.
CC SMART; SM00032; CCP; 8.
CC Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
CC Polymorphism.
CC SIGNAL 1 48
CC FT CHAIN 49 597 C4B-BINDING PROTEIN ALPHA CHAIN.
CC FT DOMAIN 49 109 SUSHI 1.
CC FT DOMAIN 112 171 SUSHI 2.
CC FT DOMAIN 174 235 SUSHI 3.
CC FT DOMAIN 238 295 SUSHI 4.
CC FT DOMAIN 298 361 SUSHI 5.
CC FT DOMAIN 364 423 SUSHI 6.
CC FT DOMAIN 425 481 SUSHI 7.
CC FT DOMAIN 483 539 SUSHI 8.
CC FT DISULFID 50 96 BY SIMILARITY.
CC FT DISULFID 81 108 BY SIMILARITY.
CC FT DISULFID 113 154 BY SIMILARITY.
CC FT DISULFID 140 170 BY SIMILARITY.
CC FT DISULFID 175 217 BY SIMILARITY.
CC FT DISULFID 203 234 BY SIMILARITY.
CC FT DISULFID 239 281 BY SIMILARITY.
CC FT DISULFID 267 294 BY SIMILARITY.
CC FT DISULFID 299 348 BY SIMILARITY.
CC FT DISULFID 332 360 BY SIMILARITY.
CC FT DISULFID 364 387 BY SIMILARITY.
CC FT DISULFID 365 387 BY SIMILARITY.
CC FT DISULFID 399 422 BY SIMILARITY.
CC FT DISULFID 426 468 BY SIMILARITY.
CC FT DISULFID 454 480 BY SIMILARITY.
CC FT DISULFID 484 525 BY SIMILARITY.
CC FT DISULFID 511 538 BY SIMILARITY.
CC FT DISULFID 546 566 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
CC FT DISULFID 558 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).

```

FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .).
FT VARIANT 92 92 Q -> T.
FT  /FTid=VAR_001977.
FT VARIANT 357 357 Y -> H.
FT  /FTid=VAR_001978.
FT VARIANT 473 473 W -> L (IN DBSNP:1801341).
FT  /FTid=VAR_012038.
SQ SEQUENCE 597 AA; 67033 MW; 67E03F2EA85A16DD CRC64;

Query Match 20.2%; Score 303; DB 1; Length 597;
Best Local Similarity 30.7%; Pred. No. 6.4e-18;
Matches 80; Conservative 42; Mismatches 101; Indels 38; Gaps 16;

Qy 20 SDQYVPEGTQAIYCRPGY-RSLGNVIMVCRK-GEWVALNPLKCRKPCGHPGDTPEGT 77
Db 67 TETREKGTTLKYCLPGYVRSHTQTLTCSNDEGW-YNTF-CIYKRCRHPGE----- 118
Qy 78 FTLTGGNV-----FEYGVKAVTCNEGYYQLLGEINRECDTD----GWTNDIPICEVVKC 128
Db 119 --LRNGQVEIKTDLFSQIEFSGEFLIGSTTSR-CEVQDRGVGWSHPLPQCEIYVKC 175
Qy 129 LPVTAPENKGIYSSAMEPDREYHFGQAVRFVCSNGYKTEGDEEMHCSDD----GFWSKEK 184
Db 176 KPPPDIRNGRHSGE-----ENFYAGFSVTSYSCDPFSLGLHASISCTVENETIGVWRPSP 231
Qy 185 PKCIVEICKSPDVINGSPISQ-KIYKENERFOYKCNNGYEYSEGRDAV--C-TESGWRP 240
Db 232 PTCEKITCRKPDVSHGEMVSGFGPIYNYKDTIVFKCQGFVL--RGSSVIHCDADSKNP 289
Qy 241 L-PSCEEKSCDN-PYIPNGDY 259
Db 290 SPPACEPNCSINLPDIPHASW 310

RESULT 5
VCP_VACCV STANDARD; PRT; 263 AA.
AC P10998;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement control protein precursor (VCP) (Secretory protein 35)
DE (Protein C3) (28 kDa protein).
GN C3L.
OS Vaccinia virus (strain WR), and
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254, 10249;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.
RC STRAIN=WR;
RX MEDLINE=88318974; PubMed=3412473;
RA Kotwal G.J., Moss B.;
RT "Vaccinia virus encodes a secretory polypeptide structurally related
RL Nature 335:176-178(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WR;
RX MEDLINE=89073756; PubMed=2849238;
RA Kotwal G.J., Moss B.;
RT "Analysis of a large cluster of nonessential genes deleted from a
RL vaccinia virus terminal transposition mutant.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=COPENHAGEN;
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
```

```

RA Paoletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [4]
RP COMPLETE GENOME.
RC STRAIN=COPENHAGEN;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL Virology 179:517-563(1990).
RN [5]
RP FUNCTION.
RX MEDLINE=92115714; PubMed=1731333;
RA Isaacs S.N., Kotwal G.J., Moss B.;
RT "Vaccinia virus complement-control protein prevents
RT antibody-dependent complement-enhanced neutralization of infectivity
RT and contributes to virulence.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:628-632(1992).
RN [6]
RP STRUCTURE BY NMR OF 146-263.
RX MEDLINE=97446168; PubMed=9299352;
RA Wiles A.P., Shaw G., Bright J., Perczel A., Campbell I.D.,
RA Barlow P.N.;
RT "NMR studies of a viral protein that mimics the regulators of
RT complement activation.";
RL J. Mol. Biol. 272:253-265(1997).
CC -I- FUNCTION: SERVES TO PROTECT THE VIRUS AGAINST COMPLEMENT ATTACK BY
CC INHIBITING BOTH CLASSICAL AND ALTERNATIVE PATHWAYS OF COMPLEMENT
CC ACTIVATION. BINDS C3B AND C4B.
CC -I- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
CC COMPLEMENT ACTIVATION (RCA).
CC -I- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; X13166; CAA31564.1; -;
DR EMBL; M22812; AAA69605.1; -;
DR EMBL; M35027; AAA47997.1; -;
DR PIR; A31005; WMVZSP.
DR PDB; 1VVC; 03-DEC-97.
DR PDB; 1VVD; 03-DEC-97.
DR PDB; 1VVE; 03-DEC-97.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
DR Signal; Repeat; Sushi; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 263 COMPLEMENT CONTROL PROTEIN.
FT DOMAIN 20 82 SUSHI 1.
FT DOMAIN 85 144 SUSHI 2.
FT DOMAIN 147 202 SUSHI 3.
FT DOMAIN 205 262 SUSHI 4.
FT DISULFID 21 70 BY SIMILARITY.
FT DISULFID 54 81 BY SIMILARITY.
FT DISULFID 86 126 BY SIMILARITY.
FT DISULFID 112 143 BY SIMILARITY.
FT DISULFID 148 190 BY SIMILARITY.
FT DISULFID 176 201 BY SIMILARITY.
FT DISULFID 206 248 BY SIMILARITY.
FT DISULFID 234 261 BY SIMILARITY.
SQ SEQUENCE 263 AA; 84322CC9A6EF8997 CRC64;

Query Match 20.1%; Score 301.5; DB 1; Length 263;
Best Local Similarity 33.3%; Pred. No. 3.4e-16;
Matches 78; Conservative 29; Mismatches 106; Indels 21; Gaps 12;
```


FT DISULFID 523 549 BY SIMILARITY.
 FT DISULFID 554 595 BY SIMILARITY.
 FT DISULFID 581 611 BY SIMILARITY.
 FT DISULFID 616 665 BY SIMILARITY.
 FT DISULFID 645 682 BY SIMILARITY.
 FT DISULFID 688 730 BY SIMILARITY.
 FT DISULFID 716 743 BY SIMILARITY.
 FT DISULFID 747 790 BY SIMILARITY.
 FT DISULFID 776 803 BY SIMILARITY.
 FT DISULFID 808 850 BY SIMILARITY.
 FT DISULFID 836 866 BY SIMILARITY.
 FT DISULFID 871 920 BY SIMILARITY.
 FT DISULFID 900 937 BY SIMILARITY.
 FT DISULFID 943 986 BY SIMILARITY.
 FT DISULFID 973 999 BY SIMILARITY.
 FT DISULFID 1004 1045 BY SIMILARITY.
 FT DISULFID 1031 1061 BY SIMILARITY.
 FT DISULFID 1066 1115 BY SIMILARITY.
 FT DISULFID 1095 1132 BY SIMILARITY.
 FT DISULFID 1138 1180 BY SIMILARITY.
 FT DISULFID 1166 1193 BY SIMILARITY.
 FT DISULFID 1197 1240 BY SIMILARITY.
 FT DISULFID 1226 1253 BY SIMILARITY.
 FT DISULFID 1258 1300 BY SIMILARITY.
 FT DISULFID 1286 1316 BY SIMILARITY.
 FT DISULFID 1321 1370 BY SIMILARITY.
 FT DISULFID 1350 1387 BY SIMILARITY.
 FT DISULFID 1396 1439 BY SIMILARITY.
 FT DISULFID 1426 1452 BY SIMILARITY.
 FT DISULFID 1457 1498 BY SIMILARITY.
 FT DISULFID 1484 1514 BY SIMILARITY.
 FT DISULFID 1519 1568 BY SIMILARITY.
 FT DISULFID 1548 1585 BY SIMILARITY.
 FT DISULFID 1591 1633 BY SIMILARITY.
 FT DISULFID 1619 1646 BY SIMILARITY.
 FT DISULFID 1650 1693 BY SIMILARITY.
 FT DISULFID 1679 1706 BY SIMILARITY.
 FT DISULFID 1711 1753 BY SIMILARITY.
 FT DISULFID 1739 1769 BY SIMILARITY.
 FT DISULFID 1774 1823 BY SIMILARITY.
 FT DISULFID 1803 1840 BY SIMILARITY.
 FT DISULFID 1848 1891 BY SIMILARITY.
 FT DISULFID 1877 1904 BY SIMILARITY.
 FT DISULFID 1909 1952 BY SIMILARITY.
 FT DISULFID 1938 1965 BY SIMILARITY.
 FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 860 860 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 959 959 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1028 1028 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1310 1310 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1481 1481 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1540 1540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1605 1605 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;

Query Match 19.9%; Score 298.5; DB 1; Length 2039;
 Best Local Similarity 27.0%; Pred. No. 6.1e-17;
 Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;

QY 3 CNE---LPPRRNTILTGSWSDQTYPEGTQAIYKRCRPGYRSLGNVINVCCKGEWVALNPL 59

Db 43 CNAPENLFPARTNLT---DEFEPICTLYNRCRPGYSGRPFISICLKNSVWTGAKD- 97
 QY 60 RKQKRCPCHPGDTPFTTTLTGNGVNFYGVKAVYTCNEGYYQLLGEINRYREC-----DTDG 115
 Db 98 -RCRRKSCRNPPDPVNGMVHVIKG--IQFGSOKYCKTGYRLIGS-SATCIIISGDTVI 153
 QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDRYHFQAVRFVCNSG-----YKLEG 168
 Db 154 WNETPICDRIPCGLPPTI-TNGDFISTNRE---NFHYGSVVYTRCNSGSGGRKKVFELVG 209
 QY 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDVINGSPIS-QKIYKENERFOYKCNMG 222
 Db 210 EPSIVCTSDNDQGVINGSPAPQCIIIPNKCTPPNVENGILVSDNRSLSLNFVVEVRCQPG 269
 QY 223 YEYSERGDVCTE-SGWRP-LPSCDE-----KSCDNP 252
 Db 270 FVMKGPVRVKQALNKWPELPSCSRVCPQPPDVLHAERTQDKDNFSPGQEVFYSCFPG 329
 QY 253 Y-----IPNGDYSP 261
 Db 330 YDLRGAASMRCTPOGDWSP 348
 RESULT 7
 C4BP_RAT
 ID C4BP_RAT STANDARD; PRT; 558 AA.
 AC Q63514;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE C4b-binding protein alpha chain precursor (C4bp).
 GN C4BPA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
 RX MEDLINE-97166082; PubMed-9013975;
 RA Hillarp A., Wiklund H., Thern A., Dahlback B.;
 RT "Molecular cloning of rat C4b binding protein alpha- and beta-chains:
 RT structural and functional relationships among human, bovine, rabbit,
 RT mouse, and rat proteins.";
 RL J. Immunol. 158:1315-1323(1997).
 CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
 CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
 CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
 CC ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
 CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
 CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
 CC AND WITH SERUM AMYLOID P COMPONENT.
 CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.
 CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
 CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC ENBL; Z50051; CAA90391.1; -;
 DR HSSP; P10998; 1VVD.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00084; sushi; 8.
 DR SMART; SM00032; CCP; 8.
 KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
 FT SIGNAL 1 13 BY SIMILARITY.
 FT CHAIN 14 558 C4B-BINDING PROTEIN ALPHA CHAIN.

63 QKRPCGHPGDTPTFTLTGCVNFYGVKAVYTCNCGYQLLGEINRECDTDG----WTN 118
 116 AKKHCNRPGLDNG--YVNGEITTFQSQIEFSCQEGFILVGS-STSSCEVRKGKGVAMSN 171
 119 DIPICEVVKLPVTPANGKIVSAMPPDREYHFGQAVRVCNSGYKIEGDEMHCS---- 175
 172 PFPECVIVKCGPPDISNGK--HSGTDFPYNHG--ISYTCDFGFLVGSFPFGC'VWN 227
 176 -DGFWSKEKPKVCIEISKSPDVINGSPIS-OKIIYKENERFOYKCNMGVEYSERGDVAC 233
 228 KTVPMVSSSPPTCEKIICSDPNLHGIVSGYKATYTHRSVRLACLNGTVLGRHRHVIC 287
 234 TEGS-WRLPSCSEKSKD-NPYIPNGDYSPL 262
 288 QGNGNWSLPTC-EFDCDLPALVNGYVTSM 317

RESULT 9
 MCP_HUMAN STANDARD; PRT; 377 AA.
 AC P15529;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Membrane cofactor protein precursor (CD46 antigen) (Trophoblast
 leucocyte common antigen) (TLX).
 GN MCP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.
 RX MEDLINE=88286080; PubMed=3260937;
 RA Lublin D.M., Liszewski M.K., Post T.W., Arce M.A., le Beau M.M.,
 RA Rebentisch M.B., Lemons R.S., Seya T., Atkinson J.P.;
 RT "Molecular cloning and chromosomal localization of the human membrane
 RT cofactor protein (MCP). Evidence for inclusion in the multigene
 RT family of complement-regulatory proteins.";
 RL J. Exp. Med. 168:181-194(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=93119658; PubMed=8418811;
 RA Cervoni F., Fenichel P., Akhoundi C., Hsi B.L., Rossi B.;
 RT "Characterization of a cDNA clone coding for human testis membrane
 RT cofactor protein (MCP, CD46).";
 RL Mol. Reprod. Dev. 34:107-113(1993).
 RN [3]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=91267562; PubMed=2050389;
 RA Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,
 RA McKenzie I.F.;
 RT "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a
 RT regulator of complement activation.";
 RL Immunogenetics 33:335-344(1991).
 RN [4]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=94014356; PubMed=7691939;
 RA Cui W., Hourcade D., Post T., Greenlund A.C., Atkinson J.P.,
 RA Kumar V.;
 RT "Characterization of the promoter region of the membrane cofactor
 RT protein (CD46) gene of the human complement system and comparison to
 RT a membrane cofactor protein-like genetic element.";
 RL J. Immunol. 151:4137-4146(1993).
 RN [5]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=92289809; PubMed=1601037;
 RA Russell S.M., Sparrow R.B., McKenzie I.F.C., Purcell D.F.J.;
 RT "Tissue-specific and allelic expression of the complement regulator
 RT CD46 is controlled by alternative splicing.";
 RL Eur. J. Immunol. 22:1513-1518(1992).
 CC !- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST

CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY
 CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD
 CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3
 CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOON WITH
 CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT
 CC ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE
 CC SYNCTIOTROPHOBLAST LAYER OF PLACENTA.
 CC !- SUBCELLULAR LOCATION: Type I membrane protein.
 CC !- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE
 CC GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN
 CC EBV-B CELLS AND LEUKEMIC CELLS. THE 66 KDa ALPHA ISOFORM AND THE
 CC 56 KDa BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND
 CC TO THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F
 CC RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME
 CC PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I
 CC AND J. SPERMATOZOON DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS
 CC 12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICE EXON 13.
 CC !- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT
 CC ERYTHROCYTES AND SOME BONE MARROW CELLS.
 CC !- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN
 CC (PROBABLE).
 CC !- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
 CC !- DATABASE: NAME=PROW; NOTE-CD guide CD46 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd46.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y00651; CAA68675.1; -
 CC EMBL: S51940; AAB24802.1; -
 CC EMBL: M58050; AAG2833.1; -
 CC EMBL: A18585; CAA01400.1; -
 CC EMBL: S65879; AAD13968.1; -
 CC PIR: S01896; S01896.
 CC HSSP: P10998; 1VVD.
 CC MIM: 120920; -
 CC InterPro: IPR000436; Sushi_SCR_CCP.
 CC Pfam: PF00084; sushi; 4.
 CC SMART: SM00032; CCP; 4.
 CC Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
 CC Sushi; Alternative splicing.
 CC SIGNAL 1 34
 CC CHAIN 35 377 MEMBRANE COFACTOR PROTEIN.
 CC DOMAIN 35 328 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 329 351 POTENTIAL.
 CC DOMAIN 352 377 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 35 95 SUSHI 1.
 CC DOMAIN 98 158 SUSHI 2.
 CC DOMAIN 161 224 SUSHI 3.
 CC DOMAIN 227 284 SUSHI 4.
 CC DOMAIN 287 311 SER/THR-RICH.
 CC DISULFID 35 80 BY SIMILARITY.
 CC DISULFID 64 94 BY SIMILARITY.
 CC DISULFID 99 141 BY SIMILARITY.
 CC DISULFID 127 157 BY SIMILARITY.
 CC DISULFID 162 210 BY SIMILARITY.
 CC DISULFID 191 223 BY SIMILARITY.
 CC DISULFID 228 270 BY SIMILARITY.
 CC DISULFID 256 283 BY SIMILARITY.
 CC CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC VARSPLIC 286 300 MISSING (IN A SECOND FORM).
 CC VARSPLIC 340 352 MISSING (IN ISOFORM M).
 CC VARSPLIC 353 361 YLSRRKKK -> DIPKGGRRKKGKQMVLMPLTRUNQPLQ
 CC QLSREA (IN ISOFORM M).
 CC TYLTDETHREVKFTSL -> KADGGAEGVATYQTKSTTPAEQ

```

FT  VARSPLIC 362 377 RG (IN ISOFORM B/D/F/H/J/L).
FT  MISSING (IN ISOFORM M AND ISOFORM N).
FT  VARSPLIC 340 361 MISSING (IN ISOFORM N).
FT  VARSPLIC 339 339 I -> ICKQVVELNMLPLRLNQLQOSREAE (IN
FT  ISOFORM N).
SQ  SEQUENCE 377 AA; 42247 MW; 2CA6F61752570B57 CRC64;

Query Match 18.3%; Score 274; DB 1; Length 377;
Best Local Similarity 29.5%; Pred. No. 9.9e-16;
Matches 77; Conservative 38; Mismatches 118; Indels 28; Gaps 13;

QY 2 DCNELPRRTEILTSQSDQTYPEGTOAIYKCRPGV---RSLGNVIMVCKGGEWALNP 58
DB 33 DACEPTTFEAMELIGK-PRPYEIGERVDYKCKGKYFIPPLATHYICDRNHTWLVSD 91
QY 59 LRKCKQPCGPGHGTPTGTTLTGNGVFYGVKAVYTCNCGYQLLG-EINRECDTQD-- 115
DB 92 -DACYRETCPYIRD-PLNGQAVPANGTYEFGYQMHFCNCGYILGIEILY--CELKGSV 147
QY 116 --WTNDIPICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVREVCNSG-----YKIEG 168
DB 148 AINSGKPPICEKVLCTPPPKIKNGKHTFSEVE---VFEYLDVATYSCDPAPGDPFSLIG 204
QY 169 DEEMHCSDGDFWSEKPKVEISCKSPDVINGSPIS---QKIIYKENERQYKCNMGY 225
DB 205 ESTIYCGDNSVMSRAAPECKVKCRFPVENGKQISGKGRFYKATVMFE--CDKGFYL 262
QY 226 SERGDVAVC-TESGW-RPLPSC 244
DB 263 DGSDTIVCDNSNTWDPVPK 283

RESULT 10
LEM2_HUMAN STANDARD; PRT; 610 AA.
AC P16581; P16111;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE (CD62E).
GN SELE OR ELAM1.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90175355; PubMed-1689848;
RA Hession C., Osborn L., Goff D., Chi-Rosso G., Vassallo C.,
RA Fasek M., Pittack C., Tizard R., Goelz S., McCarthy K., Hopple S.,
RA Lobb R.;
RT "Endothelial leukocyte adhesion molecule 1: direct expression cloning
RT and functional interactions.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1673-1677(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-89162047; PubMed-2466335;
RA Bevilacqua M.P., Stengelin S., Gimbrone M.A. Jr., Seed B.;
RT "Endothelial leukocyte adhesion molecule 1: an inducible receptor for
RT neutrophils related to complement regulatory proteins and lectins.";
RL Science 243:1160-1165(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-91115870; PubMed-1703529;
RA Collins T., Williams A., Johnston G.I., Kim J., Eddy R., Shows T.,
RA Gimbrone M.A. Jr., Bevilacqua M.P.;
RT "Structure and chromosomal location of the gene for endothelial-
RT leukocyte adhesion molecule 1.";
RL J. Biol. Chem. 266:2466-2473(1991).
RN [4]

```

```

RP LIGAND.
RX MEDLINE-91068005; PubMed-1701274;
RA Phillips M.L., Nudelman E., Gaeta F.C., Perez M., Singhal A.K.,
RA Hakomori S., Paulson J.C.;
RT "ELAM-1 mediates cell adhesion by recognition of a carbohydrate
RT ligand, sialyl-Lex.";
RL Science 250:1130-1132(1990).
RN [5]
RP 3D-STRUCTURE MODELING OF LECTIN DOMAIN.
RX MEDLINE-93202275; PubMed-7681016;
RA Mills A.;
RT "Modelling the carbohydrate recognition domain of human E-selectin.";
RL FEBS Lett. 319:5-11(1993).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 22-178.
RX MEDLINE-94150646; PubMed-7509040;
RA Graves B.J., Crowther R.L., Chandran C., Rumberger J.M., Li S.,
RA Huang K.-S., Presky D.H., Familletti P.C., Wolitzky B.A., Burns D.K.;
RT "Insight into E-selectin/ligand interaction from the crystal
RT structure and mutagenesis of the lec/EGF domains.";
RN Nature 367:532-538(1994).
RP VARIANT ARG-149.
RX MEDLINE-95179107; PubMed-7533025;
RA Wenzel K., Felix S., Kleber F.X., Brachold R., Menke T., Schattke S.,
RA Schulte K.L., Glaser C., Rohde K., Baumann G., Speer A.;
RT "E-selectin polymorphism and atherosclerosis: an association study.";
RL Hum. Mol. Genet. 3:1935-1937(1994).
RN [8]
RP VARIANT ARG-149.
RX MEDLINE-99134508; PubMed-9933738;
RA Ye S.-Q., Usher D., Virgil D., Zhang L.Q., Yochim S.E., Gupta R.;
RT "A PstI polymorphism detects the mutation of serine-128 to arginine in
RT CD 62E gene - a risk factor for coronary artery disease.";
RL J. Biomed. Sci. 6:18-21(1999).
CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC ELAM-1 IS SIALYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF
CC POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC GLYCOLIPIDS).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- POLYMORPHISM: A POLYMORPHISM IN POSITION 149 IS ASSOCIATED WITH A
CC HIGHER RISK OF CORONARY ARTERY DISEASE (CAD). A SIGNIFICANTLY
CC HIGHER MUTATION FREQUENCY (ARG-149) IS OBSERVED IN PATIENTS WITH
CC ANGIOGRAPHICALLY PROVEN SEVERE ATHEROSCLEROSIS COMPARED WITH AN
CC UNSELECTED POPULATION (SER-149).
CC -1- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD62E entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd62e.htm".
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M30640; AAA52377.1; -
DR EMBL; M61893; AAA52375.1; -
DR EMBL; M61895; AAA52375.1; JOINED.
DR EMBL; M61887; AAA52375.1; JOINED.
DR EMBL; M61898; AAA52375.1; JOINED.
DR EMBL; M61890; AAA52375.1; JOINED.
DR EMBL; M61891; AAA52375.1; JOINED.
DR EMBL; M61892; AAA52375.1; JOINED.
DR EMBL; M24736; AAA52376.1; -
DR PIR; A32606; A32606.
DR PIR; A35046; A35046.
DR PIR; A38615; A38615.

```

```
DR PDB; 1ESL; 31-AUG-94.
DR PDB; 1KJA; 03-APR-96.
DR MIM; 131210; .
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002396; Selection.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 6.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 6.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE-LECTIN_2; 1.
KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW Selectin; Signal; Sushi; Repeat; Polymorphism; 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 610
FT DOMAIN 22 556
FT TRANSEM 557 578
FT DOMAIN 579 610
FT DOMAIN 38 138
FT DOMAIN 139 175
FT DOMAIN 179 238
FT DOMAIN 241 300
FT DOMAIN 303 363
FT DOMAIN 366 426
FT DOMAIN 429 489
FT DOMAIN 492 548
FT DISULFID 40 138
FT DISULFID 111 130
FT DISULFID 143 154
FT DISULFID 148 163
FT DISULFID 165 174
FT DISULFID 180 224
FT DISULFID 210 237
FT DISULFID 242 286
FT DISULFID 272 299
FT DISULFID 304 349
FT DISULFID 335 362
FT DISULFID 367 412
FT DISULFID 398 425
FT DISULFID 430 475
FT DISULFID 461 488
FT DISULFID 493 534
FT DISULFID 520 547
FT CARBOHYD 25 25
FT CARBOHYD 145 145
FT CARBOHYD 160 160
FT CARBOHYD 179 179
FT CARBOHYD 199 199
FT CARBOHYD 203 203
FT CARBOHYD 265 265
FT CARBOHYD 312 312
FT CARBOHYD 332 332
FT CARBOHYD 503 503
FT CARBOHYD 527 527
FT VARIANT 130 130
FT VARIANT 149 149
FT VARIANT 295 295
FT VARIANT 421 421
FT VARIANT 468 468
FT VARIANT 575 575
/FTid=VAR_011794.
SQ SEQUENCE 610 AA; 66655 MW; 7D43E3CD01229229 CRC64;
Query Match 18.2%; Score 272; DB 1; Length 610;
Best Local Similarity 27.1%; Pred. No. 2.5e-15;
Matches 70; Conservative 37; Mismatches 107; Indels 44; Gaps 15;
QY 23 TYPESTQAIYKCRPGYRSLGNVIMVC-RKGEWVALNPLRK---CQ--KRP-----CGHP 70
DB 261 SFPWNTTCTFDCEEGFELMGAQSLQCTSSGNWMDNEKPTCKAVTCRAVRPQNGSVRCSSH- 319
QY 71 GDTPTGFTTLTGNNVFYGVKAVYTCNEGYYOLLGEINRYECDTDG-WTNDIPICEVVKACL 129
DB 320 --SPAGEFT-----FKSSCNFTCEGFMQLQPAQV-ECTTQGTQWTOQIIPVCEAFQCT 368
QY 130 PVTAPENGKIVSSAMEPDREYHFQAVRVFVNSGYKIEGDEMHCSDDGFWSKPKKPCVCE 189
DB 369 ALSNPERG-YMNCPLPSASGSFRYGSCEFSCEQGFVLKSGKRLQCGPTGEMDNKPTCEA 427
QY 190 ISCKSPDVIN-----GSPISOKIYKENERFOYKCNMGYSEYSERGDAVCTESC-W 238
DB 428 VRC---DAVHQPPLGLVRCASHPIGE-FTYKSSCAF--SCBEGFELHGSTQLECTSQSGW 481
QY 239 -RPLPSCSEKSCDNPYIP 255
DB 482 TEEVPSQGVKCSSLAVP 499
RESULT 11
LEM2_PIG STANDARD; PRT; 484 AA.
AC P98110;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE (CD62E).
GN SELE.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Aortic endothelium;
RX MEDLINE=95071392; PubMed=7526854;
RA Rollins S.A., Evans M.J., Johnson K.K., Elliot E.A., Squinto S.P.,
RA Matis L.A., Rother R.P.;
RA "Molecular and functional analysis of porcine E-selectin reveals a
RT potential role in xenograft rejection.";
RL Biochem. Biophys. Res. Commun. 204:763-771(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Aortic endothelium;
RX MEDLINE=94271236; PubMed=7516159;
RA Tsang Y.T.M., Haskard D.O., Robinson M.K.;
RA "Cloning and expression kinetics of porcine vascular cell adhesion
RT molecule.";
RL Biochem. Biophys. Res. Commun. 201:805-805(1994).
CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC ELAM-1 IS SIALLYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF
CC POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC GLYCOLIPIDS).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN ACUTE CELLULAR ALLOGRAFT
CC REJECTION AND PROBABLY ALSO IN XENOGRAFT REJECTION.
CC -1- SURCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS; PORCINE E-LECTIN LACKS
```


FT DOMAIN 173 235 SUSHI 3.
FT DOMAIN 238 295 SUSHI 4.
FT DOMAIN 298 363 SUSHI 5.
FT DOMAIN 366 426 SUSHI 6.
FT DOMAIN 428 484 SUSHI 7.
FT DOMAIN 486 542 SUSHI 8.
FT DISULFID 50 95 BY SIMILARITY.
FT DISULFID 80 107 BY SIMILARITY.
FT DISULFID 112 153 BY SIMILARITY.
FT DISULFID 139 169 BY SIMILARITY.
FT DISULFID 174 217 BY SIMILARITY.
FT DISULFID 203 234 BY SIMILARITY.
FT DISULFID 239 281 BY SIMILARITY.
FT DISULFID 267 294 BY SIMILARITY.
FT DISULFID 299 350 BY SIMILARITY.
FT DISULFID 334 362 BY SIMILARITY.
FT DISULFID 366 390 BY SIMILARITY.
FT DISULFID 367 412 BY SIMILARITY.
FT DISULFID 402 425 BY SIMILARITY.
FT DISULFID 429 471 BY SIMILARITY.
FT DISULFID 457 483 BY SIMILARITY.
FT DISULFID 487 528 BY SIMILARITY.
FT DISULFID 514 541 BY SIMILARITY.
FT DISULFID 549 561 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 561 561 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 525 525 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 610 AA; 68886 MW; D806B270E8A06B58 CRC64;

Query Match 17.7%; Score 265.5; DB 1; Length 610;
Best Local Similarity 27.5%; Pred. No. 8.7e-15;
Matches 76; Conservative 45; Mismatches 110; Indels 45; Gaps 16;

QY 6 LPRRTEIITLGSWSDQTYPEGTQAIKCRPGYR--SLGNVIMVCRKGEWALNLRKQC 63
DB 52 IIPYDFAFPINELNETRFETGTLRYTCRPGYRISRNKFLICDGTDNW---KYREFCV 108
QY 64 KRPCGHGPDTPFTTITGNV-----FEYGVRAVTCNEGYYOLLGEIN-YRECDTDG-- 115
DB 109 KRRCENPGE-----LLNGQVIVKTDYSFGSIEFSCSEGYVLIGSANSYCOLQDKGVV 161
QY 116 WTNDIFICEVVKLPVAPENGKIVSAMEPDREYHFGQAVREVCNSGYKIEGDEMHCS 175
DB 162 WSDPLPQCIACKEPPPTTISNGR--HNGGDED-FYTYGSSVTYSCDRDFSMCLKASISCR 218
QY 176 DD----GFSWKEKPKVCISKSPDVINGSPISQKI-----IYKENERFOYKCNMGYEV 225
DB 219 VENKTIQVNSPSPCKKVICVOPVYKDG-----KITSGFGPIITYQOSIVYACNKGFR 273
QY 226 SERGDV--C-TESGWR-PLPSCSEKSC-DNPYIPN 256
DB 274 --EGDSLIIHCADSNWNPPTCELNGCLGLPHIP 307

RESULT 13
CCPH_HSVSA
ID CCPH_HSVSA STANDARD; PRT; 360 AA.
AC Q01016;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DT Complement control protein homolog precursor (CCPH).
GN 4 OR CCPH.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;

RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
RN [2]
RP SIMILARITY TO CCP.
RX MEDLINE=92260674; PubMed=1316492;
RA Albrecht J.-C., Fleckenstein B.;
RT "New member of the multigene family of complement control proteins in
herpesvirus saimiri.";
RL J. Virol. 66:3937-3940(1992).
CC -!- ALTERNATIVE PRODUCTS: A MEMBRANE-BOUND FORM AND A SECRETED FORM
ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC -!- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
COMPLEMENT ACTIVATION (RCA).
CC -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64346; CAA45626.1; -;
DR EMBL; X64346; CAA45627.1; -;
DR EMBL; X60283; CAA42823.1; -;
DR EMBL; X60283; CAA42822.1; -;
DR PIR; B42534; WMBE2E.
DR PIR; A42534; WMBE1E.
DR PIR; S24567; S24567.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
KW Signal; Repeat; SUSHI; Transmembrane; Alternative splicing;
KW Glycoprotein.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 360 COMPLEMENT CONTROL PROTEIN HOMOLOG.
FT DOMAIN 83 143 SUSHI 1.
FT DOMAIN 146 206 SUSHI 2.
FT DOMAIN 209 265 SUSHI 3.
FT TRANSMEM 328 350 POTENTIAL.
FT DISULFID 84 125 BY SIMILARITY.
FT DISULFID 111 142 BY SIMILARITY.
FT DISULFID 147 191 BY SIMILARITY.
FT DISULFID 175 205 BY SIMILARITY.
FT DISULFID 210 252 BY SIMILARITY.
FT DISULFID 238 264 BY SIMILARITY.
FT CARBOHYD 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 289 302 RICNGNCTSMPTQ -> AECACPGSNYPIS (IN
SHORT ISOFORM).
FT VARSPLIC 303 360 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 360 AA; 40006 MW; 6278A6C2ECD49669 CRC64;

Query Match 17.7%; Score 264.5; DB 1; Length 360;
Best Local Similarity 30.6%; Pred. No. 5.8e-15;
Matches 76; Conservative 22; Mismatches 119; Indels 31; Gaps 11;

QY 24 YPEGTOAIYKCRPGYRSLGNVIMVCRKGEWALNLRKQKRCQKPGHGPDTPTFTLTGG 83
DB 44 YPNGTTLHVTCTREGYAKRPVQVTCVNGWTV---PKKCKKKKCSPTQDLLNGRYTIVT-G 99
QY 84 NVFEGVGVKAVYTCNEGYYQLLGEIN-----YRECDTDGWTNDIPICEVVKLPVAPENGKI 139
DB 139 NVFEGVGVKAVYTCNEGYYQLLGEIN-----YRECDTDGWTNDIPICEVVKLPVAPENGKI 139

```

Db 100 NLX-YGSVITTCNSGYSLIGSTTSACLLKRGGRVDWTPRPPICDIKKCKPPQIANG-- 156
QY 140 VSSAMEPDREYHFGQAVRFVCNSGYK--IRGDEMHCSDDGFV-SREKPKVCVEISCKSPD 196
Db 157 --THTNWKDYTYLDTVYSCNDETCLTLTGSPSSKLCSETGSGWPNGETKCEFIKCLPQ 214
QY 197 VINGSPISQKIYKENERFOY-----KCNMGYSYSERGDAVCTESQWRP-LPSCSEKSCD 250
Db 215 VAN----AYVEVRKSATSMQYLHINVKYKGFMLYGETPNTCNHGYWSPAIPCEMKISS- 269
QY 251 NPVIPNGD 258
Db 270 ----PKGD 273

RESULT 14
DAF_HUMAN
ID DAF_HUMAN STANDARD; PRT; 381 AA.
AC P08174; P09679; P78361;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Complement decay-accelerating factor precursor (CD55 antigen).
GN DAF OR CR OR CD55.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=87115845; PubMed=2433596;
RA Caras I.W., Davitz M.A., Rhee L., Weddell G., Martin D.W. Jr.,
RA Nussenzweig V.;
RT "Cloning of decay-accelerating factor suggests novel use of splicing
RL to generate two proteins.";
RN Nature 325:545-549(1987).
RP [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 6-381 FROM N.A. (ISOFORM 2).
RX MEDLINE=87175602; PubMed=2436222;
RA Medof M.E., Lublin D.M., Hollers V.M., Ayers D.J., Getty R.R.,
RA Leykam J.F., Atkinson J.P., Tykocinski M.L.;
RT "Cloning and characterization of cDNAs encoding the complete sequence
RL of decay-accelerating factor of human complement.";
RN Proc. Natl. Acad. Sci. U.S.A. 84:2007-2011(1987).
RP [4]
RP SEQUENCE OF 35-381 FROM N.A. (ISOFORM 2).
RC TISSUE=Hippocampus;
RA Kumar V.B., Hyung C., Nakra R., Walters M., Sasser T., Bernardo A.;
RT "Decay-accelerating factor (DAF; CD 55) in the brain of Alzheimer's
RL disease patients";
RN Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RP [5]
RP SEQUENCE OF 1-100 FROM N.A.
RX MEDLINE=91271256; PubMed=1711208;
RA Ewulonu U.K., Ravi L., Medof M.E.;
RT "Characterization of the decay-accelerating factor gene promoter
RL region.";
RN Proc. Natl. Acad. Sci. U.S.A. 88:4675-4679(1991).
RP [6]
RP SEQUENCE OF 35-46.
RC TISSUE=Urine;
RX MEDLINE=91291869; PubMed=1712233;
RA Nakano Y., Suigita Y., Ishikawa Y., Choi N.-H., Tobe T., Tomita M.;
RT "Isolation of two forms of decay-accelerating factor (DAF) from human
RL urine.";
RN Biochim. Biophys. Acta 1074:326-330(1991).
RP [7]
RP GPI-ANCHOR.

```

```

RX MEDLINE=91093238; PubMed=1824699;
RA Moran P., Raab H., Kohr W.J., Caras I.W.;
RT "Glycophospholipid membrane anchor attachment. Molecular analysis of
RL the cleavage/attachment site.";
RN J. Biol. Chem. 266:1250-1257(1991).
RP [8]
RP DISULFIDE BONDS IN SUSHI DOMAINS.
RX MEDLINE=92305034; PubMed=1377029;
RA Nakano Y., Sumida K., Kikuta N., Miura N.-H., Tobe T., Tomita M.;
RT "Complete determination of disulfide bonds localized within the short
RL consensus repeat units of decay accelerating factor (CD55 antigen).";
RN Biochim. Biophys. Acta 1116:235-240(1992).
RP [9]
RP FUNCTION AS A ECHOVIRUS RECEPTOR.
RX MEDLINE=95045399; PubMed=7525274;
RA Ward T., Pipkin P.A., Clarkson N.A., Stone D.M., Minor P.D.,
RA Almond J.W.;
RT "Decay-accelerating factor CD55 is identified as the receptor for
RL echovirus 7 using CELICS, a rapid immuno-focal cloning method.";
RN EMBO J. 13:5070-5074(1994).
RP [10]
RP VARIANT BLOOD GROUP DR(A-).
RX MEDLINE=94325573; PubMed=7519480;
RA Lublin D.M., Wallinson G., Poole J., Reid M.E., Thompson E.S.,
RA Ferdinand B.R., Telen M.J., Anstee D.J., Tanner M.J.A.;
RT "Molecular basis of reduced or absent expression of
RL decay-accelerating factor in Cromer blood group phenotypes.";
RN Blood 84:1276-1282(1994).
RP [1]
RP FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES
CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND
CC FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS
CC THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF
CC THE COMPLEMENT CASCADE.
RP [1]
RP FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECHOVIRUS 7 AND RELATED
CC VIRUSES (ECHOVIRUSES 13, 21, 29 AND 33).
RP [1]
RP SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT
CC HOMODIMER (MINOR FORM)
RP [1]
RP SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC HERE; ARE PRODUCED BY ALTERNATIVE SPLICING.
RP [1]
RP TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL
CC TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA MEMBRANES OF ALL CELL
CC PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS
CC LINING EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE
CC ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.
RP [1]
RP DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC ACTIVE SITE ON SCR3 (BY SIMILARITY).
RP [1]
RP PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.
RP [1]
RP POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CROMER BLOOD GROUP
CC SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CR(A),
CC TC(A), DR(A), ES(A), WES(B), UMC, AND IFC) AND LOW-INCIDENCE
CC (TC(B), TC(C), AND WES(A)) ANTIGENS THAT RESIDE ON DAF. IN THE
CC CROMER PHENOTYPES DR(A-) AND INAB THERE IS REDUCED OR ABSENT
CC EXPRESSION OF DAF, RESPECTIVELY. IN THE CASE OF THE DR(A-)
CC PHENOTYPE, A SINGLE NUCLEOTIDE SUBSTITUTION WITHIN EXON 5 ACCOUNTS
CC FOR TWO CHANGES: A SIMPLE AMINO ACID SUBSTITUTION THAT IS THE
CC BASIS OF THE ANTIGENIC VARIATION, AND AN ALTERNATIVE SPLICING
CC EVENT THAT UNDERLIES THE DECREASED EXPRESSION OF DAF IN THIS
CC PHENOTYPE.
RP [1]
RP SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
RP [1]
RP SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC (RCA) FAMILY.
RP [1]
RP DATABASE: NAME=PROW; NOTE=CD guide CD55 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cg/cd55.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2002, 06:16:26 ; Search time 77.9 Seconds
(without alignments)
588.494 Million cell updates/sec

Title: US-09-316-163-10
Perfect score: 1497
Sequence: 1 EDCNELPRRNTTEILGWSWS.....EKSCDNPYPNGDYSPLRIK 265

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1497	100.0	1172	Q9NU87	Q9nu87 homo sapien
2	1090	72.8	1236	Q91YB6	Q91yb6 rattus norv
3	699	46.7	669	Q28085	Q28085 bos taurus
4	440	29.4	1053	Q91275	Q91275 paralabrax
5	332.5	22.2	360	Q9YVTO8	Q9yvtq8 ateline her
6	322.5	21.5	550	Q288903	P88903 kaposi's sa
7	322.5	21.5	550	Q40912	O40912 kaposi's sa
8	320	21.4	645	Q9WRU2	Q9wru2 macaca mula
9	310	20.7	259	Q287616	P87616 cowpox viru
10	308.5	20.6	395	Q9J2M6	Q9j2m6 macaca mula
11	308.5	20.6	522	Q28769	Q28769 papio cynoc
12	301	20.1	679	Q99254	Q99254 mus musculu
13	300.5	20.1	3567	Q9ES77	Q9es77 mus musculu
14	298.5	19.9	559	Q9UQV2	Q9uqv2 homo sapien
15	298.5	19.9	2039	Q16745	Q16745 homo sapien
16	298.5	19.9	2489	Q16744	Q16744 homo sapien

ALIGNMENTS

RESULT 1

Q9NU87 PRELIMINARY; PRT; 1172 AA.

AC Q9NU87; DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE DJL77P10.1.1 (H FACTOR 1 (COMPLEMENT) ISOFORM 1).

GN HFI.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Bird C.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL049744; CAB70597.1; -

DR HSSP; P086603; 1HFH.

DR InterPro; IPR000436; SushL_SCR_CCP.

DR Pfam; PF00084; sushl; 19.

DR SMART; SM00032; CCP; 19.

SQ SEQUENCE 1172 AA; 132087 MW; 8F5B954C4B4FA454 CRC64;

Query Match 100.0%; Score 1497; DB 4; Length 1172;
Best Local Similarity 100.0%; Pred. No. 6.9e-135;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCNELPRRNTTEILGWSWSDDTYPEGTQAIYKCRGYSLGNVIMVCRGEWALNPLR 60
|||||

Db 19 EDCNELPRRNTTEILGWSWSDDTYPEGTQAIYKCRGYSLGNVIMVCRGEWALNPLR 78
|||||

QY 61 KCQKPCGHPGDTPTFTLTGNGVFEYGVKAVYTCNEGYYQLLGEINRECDTGTNDI 120
|||||

Db 79 KCQKPCGHPGDTPTFTLTGNGVFEYGVKAVYTCNEGYYQLLGEINRECDTGTNDI 138
|||||

QY 121 PTCEVVKLPVTAPENKTKTVSSAMEPDRYHFQGAQVRFVNCVSGYKTEGDEMHCSDDGF 180
|||||

Db 139 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVNSGYKIEGDEEMHCSDDGF 198

QY 181 SKKPKCVETSCSPDVINGSPISQIIYKENERFOYKCNMGVEYSESGDAVCTESGWRP 240

Db 199 SKKPKCVETSCSPDVINGSPISQIIYKENERFOYKCNMGVEYSESGDAVCTESGWRP 258

QY 241 LPSCEKSCDNPYPNGDYSPLRIK 265

Db 259 LPSCEKSCDNPYPNGDYSPLRIK 283

RESULT 2

Q91YB6 PRELIMINARY; PRT; 1236 AA.

AC Q91YB6;

DT 01-DEC-2001 (T-EMBLrel. 19, Created)

DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)

DE COMPLEMENT INHIBITORY FACTOR H.

GN FH.

OS Rattus norvegicus (Rat).

OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;

RA Demberg T., Goetze O., Schlaf G.;

RT "Rat complement factor H: molecular cloning, sequencing and expression

RT in tissues and isolated cells.";

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

QY EMBL; AJ320522; CAC67513.1;

Db SEQUENCE 1236 AA; 140343 MW; 1AC89PFA28232EBF CRC64;

Query Match 72.8%; Score 1090; DB 11; Length 1236;

Best Local Similarity 71.3%; Pred. No. 8.4e-96;

Matches 189; Conservative 25; Mismatches 51; Indels 0; Gaps 0;

QY 1 EDCNELPPRNTTEILTGSWSDQTYPECTQAIYKCRPGYRSLGNVIMVCRKGEMVALNPLR 60

Db 19 EDCGGPPPRENSELSSWSSEQLYSECTQATYKCRPGYRSLGTIVKCKNGEWPSPSR 78

QY 61 KCKRCPCHPGDTPFGFTTGTGNVFYKAVYTCNEGQYLLGEINRECDTDGWTNDI 120

Db 79 ICKRCPCHPGDTPFGFRLAVGSEFEFGAKVYTCNEGQYLLGEIDYRECDADGWTNDI 138

QY 121 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVNSGYKIEGDEEMHCSDDGF 180

Db 139 PICEVVKCLPVTLENGRIVSGAEPDQYIFGQVRFECNSGFKIBGQKEMHCSENGLW 198

QY 181 SKKPKCVETSCSPDVINGSPISQIIYKENERFOYKCNMGVEYSESGDAVCTESGWRP 240

Db 199 SNEKPCQVEISCLPPRVENGDIYLPVYKENERFOYKCKQGFYKRGDAVCTGSGWNP 258

QY 241 LPSCEKSCDNPYPNGDYSPLRIK 265

Db 259 QPSCCEMTCLTPYIPNGIYTPHRIK 283

RESULT 3

Q28085 PRELIMINARY; PRT; 669 AA.

AC Q28085;

DT 01-NOV-1996 (T-EMBLrel. 01, Created)

DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)

DE CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RP SEQUENCE FROM N.A.

RC TISSUE=LIVER;

RX MEDLINE-96202005; PubMed-8615824;

RA Soames C.J., Day A.J., Sim R.B.;

RT "Prediction from sequence comparisons of residues of factor H involved

RT in the interaction with complement component C3b.";

RL Biochem. J. 315:523-531(1996).

DR EMBL; X98697; CAA67257.1;

DR HSP; P10998; LVVD.

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam; PF00084; sushi; 11.

DR SMART; SM00032; CCP; 11.

FT NON_TER 1

FT NON_TER 669

FT NON_TER 669

SQ SEQUENCE 669 AA; 75683 MW; D0D9DB30EE747AC2 CRC64;

Query Match 46.7%; Score 699; DB 6; Length 669;

Best Local Similarity 63.4%; Pred. No. 1.3e-58;

Matches 118; Conservative 26; Mismatches 42; Indels 0; Gaps 0;

QY 76 GTFTLTGNNVFYKAVYTCNEGQYLLGEINRECDTDGWTNDIPICEVVKCLPVTAPE 135

Db 3 GSPLAEGNQFEYKAVVYTCDEGYQMVGMNFRECDTNGWTNDIPICEVVKCLPVTPE 62

QY 136 NGKIVSSAMEPDREYHFGQAVRFVNSGYKIEGDEEMHCSDDGFWSKEKPKCVETSCSP 195

Db 63 NGKIFSDALEPDQBYTYTQVVFECNSGYMLDGPQIHCSAGGVMSAETPKCVETFCPP 122

QY 196 DVINGSPIQIIYKENERFOYKCNMGVEYSESGDAVCTESGWRPLPSCSEKSCDNPV 255

Db 123 VILNGQAVLPKATYKQNERVQYRCAGFEYRGORDYCTKSGWTPATCIBITCDPPRP 182

QY 256 NGDYSP 261

Db 183 NGVYRP 188

RESULT 4

Q91275 PRELIMINARY; PRT; 1053 AA.

AC Q91275;

DT 01-NOV-1996 (T-EMBLrel. 01, Created)

DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)

DE COMPLEMENT REGULATORY PLASMA PROTEIN.

OS Parabrax nebulifer (barred sand bass).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;

OC Serranidae; Paralabrax.

OX NCBI_TaxID=30873;

[1]

RP SEQUENCE FROM N.A.

RC TISSUE=LIVER;

RX MEDLINE-94318039; PubMed-8042982;

RA Dahmen A., Kaldon T., Zipfel P.F., Gigli I.;

RT "Cloning and characterization of a cDNA representing a putative

RT complement-regulatory plasma protein from barred sand bass (Parablax

RT nebulifer).";

RL Biochem. J. 301:391-397(1994).

DR EMBL; L21703; AAA92556.1;

DR HSP; P08603; LHFH.

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam; PF00084; sushi; 16.

DR SMART; SM00032; CCP; 16.

SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D5D3 CRC64;

Query Match 29.4%; Score 440; DB 13; Length 1053;

Best Local Similarity 36.2%; Pred. No. 1.6e-33;

Matches 89; Conservative 35; Mismatches 108; Indels 14; Gaps 6;

```
QY 21 DQTPGEGTOAIYKCRGYSRLGNVIMVCRKGEWALNPLRKKOKRPGCHPGDTPFGFTTL 80
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
47 EASYPGGROVRVCGNVGYS--GFFKLVCVEGKWETRG--AKQCPSCGHPGDAQADFLH 102
QY 81 TGGNVEFYGVKAVYTCNEGVLGELNRECDTGDWNTDIPICEVVKCLPVTAPENGKIV 140
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
103 AEGNDFVFGSVYTCOKGQVMYSRINRYRCVAEGWGVVPCESQOCPLIHVDNNVQVI 162
QY 141 SSAMEPDREYHFGQAVRVCNSGYKI-EGDEEMHCSDDGFWSEKPKVCVEISCKSPDVIN 199
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
163 GG---PEATGCVNRFSCSKSEILDGSPELYCDERGDWSPGPKCAITCAIPIEN 218
QY 200 GSPISOKIYKENERFOYKCNMGYEYSERGDAVCTESG----WRPLPSCSEKSCDNPYP 255
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
219 GNPVGAIREYKENDVLHYECDRAFKHIDR--PSTCIKOGIKAEMSP7PLCESIKRLTIMD 277
QY 256 NGDYSP 261
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
278 GTRYEP 283

RESULT 5
QYTO8 PRELIMINARY; PRT; 360 AA.
AC Q9YTO8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE COMPLEMENT CONTROL PROTEIN HOMOLOG CCPH.
OS AteLine herpesvirus 3.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=85618;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=73;
RX MEDLINE=200911363; PubMed=10623770;
RA Albrecht J.C.;
RT "Primary structure of the Herpesvirus Ateles genome.";
RL J. Virol. 74:1033-1037(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=73;
RA Albrecht J.-C., Fleckenstein B.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083424; AAC95530.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
SQ SEQUENCE 360 AA; 40208 MW; 118CF83C034352A0 CRC64;

Query Match 22.2%; Score 332.5; DB 12; Length 360;
Best Local Similarity 34.9%; Pred. No. 9,1e-24;
Matches 96; Conservative 32; Mismatches 106; Indels 41; Gaps 18;

QY 8 PRNTEI-----LTGSWSDQIYPGTOAIYKCRPGYSRLGNVIMVCRKGEWALNPL 59
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
24 PKRNRVSLRYSLRYNITN--SSGSYPNGTTLQVTCRKGY--IGRQIQTVTCVNGNWTVPN-- 77
QY 60 RKCQRKPCGHPGDTFGTFTLGGNVFEYGVKAVYTCNEGYSRLGNVIMVCRKGEWALNPL 59
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
78 -EQARRKSTPADLLNGWTVT-GNLY-VGSVITYTCNTGYQLLGSP7-SSCLLGPDRGV 133
QY 116 -WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYH-FGQAVRVCNSGYK--TEGDEE 171
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
134 NWTPRPICEITKCRPPPTPIANGTHNI-----KEYTYLDVAVTYSNDETKLTGTGSS 188
QY 172 MHCSDGFW-SKEPKVCVEISCKSPDVINSGSPISOKIYKENERFOY---KCNMGYEYSE 227
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
189 KQCSETGRWPFDEETKCFKVKIPQVANGHVEVRKT--SNNVQYQYVNIKCDKGFRLQG 246
```

```
QY 228 RGDVACTESGWRP-LPSCSEK---CDNPYIPNGD 258
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
247 ETPNMCKNGWFFPALPTCEKPAAPPGRGMDPHDSGE 281

RESULT 6
P88903 PRELIMINARY; PRT; 550 AA.
ID P88903;
AC 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF 4.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=37296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97094384; PubMed=8939871;
RA Moore P.S., Boshoff C., Weiss R.A., Chang Y.;
RT "Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV.";
RL Science 274:1739-1744(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97121480; PubMed=8962146;
RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D., Party J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RT "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8).";
RL Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D., Party J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U75698; AAC57082.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
SQ SEQUENCE 550 AA; 60688 MW; D4B8B2B4BACD1CB5 CRC64;

Query Match 21.5%; Score 322.5; DB 12; Length 550;
Best Local Similarity 31.2%; Pred. No. 1.4e-22;
Matches 72; Conservative 31; Mismatches 109; Indels 19; Gaps 9;

QY 33 KCRPGYSRLG-NVIMVC-RKGEWALNPLRKKOKRPGCHPGDTPFGTFTLTGG-NVFEY 89
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
52 RCRSGYTVARNITATCLOGGTW--SEPTATCNKSCPNPGEIQNGKVIHGGQDALKYG 109
QY 90 VRAYVTCNEGYSRLGNVIMVCRKGEWALNPLRKKOKRPGCHPGDTPFGTFTLTGG-NVFEY 144
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
110 ANISVYCNVEYFLVGYRYVRYCMIGASGQMAWSSSPFCEKEKC-----HRPKINGDF 163
QY 145 EPDREYH-FGQAVRVCNSGYKIEGDEEMHCSDDGFWSEKPKVCVEISCKSPDVINSGPI 203
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
164 KPDKDYEYNDVAVHFECEGYSRLGNVIMVCRKGEWALNPLRKKOKRPGCHPGDTPFGTFTLTGG 223
QY 204 SQ-KIYKENERFOYKCNMGYEYSERGDAVCTESG--RPLPSCSEKSCDNP 252
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
224 QGFSLTYYKHQSVTACNDGFLVRLGSP7ITTCNVNTWDPPLPKCVLEDIDDP 274

RESULT 7
O40912 PRELIMINARY; PRT; 550 AA.
ID O40912;
AC O40912;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
```



```
Query Match      20.7%; Score 310; DB 12; Length 259;
Best Local Similarity 33.5%; Pred. No. 8.7e-22;
Matches 84; Conservative 29; Mismatches 116; Indels 22; Gaps 13;

QY 3 CNELPPRRNTEILTGSGWSQDTYPEGTQAIYKCRPGYR--SLGNVIMVCRKGEWALNPLR 60
DB 20 CCPIPSRPITMKFGT-VDSHNIGDTIYELCPGKOKMGPIYAKCTGTGWTLEN--- 75

QY 61 KCQKRCPCGHPGDTPTGFTLTGCVNFEYGVKAVYTCNKGQYLLGEI-NYRECDTDG---W 116
DB 76 QCIKRRCPSPRDINDQDLGG---VDFGSSITYSCNSGYHLIGESKSYCELGSTGSMW 132

QY 117 TNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRVCNCGYKIEGDEEMHCS 176
DB 133 NEPAFICESVKQSPPSISNGR--HNGYED--FYTDGSVVITYSCNSGYSLIGNSGYLCS- 187

QY 177 DGFWSKEPKVCVEISCKSPDVINGS-PISQKIYKENERFQYKCNMGYYSERGDVCTE 235
DB 188 GGEWS-PPPTCOIVKCPHTTISNGYLSGFKRSYNDNVDFKCKGYKLSSSSSSTCSP 246

QY 236 SG-WRP-LPSC 244
DB 247 GNTWQPELPKC 257

RESULT 10
ID Q9J2M6 PRELIMINARY; PRT; 395 AA.
AC Q9J2M6;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE COMPLEMENT BINDING PROTEIN.
OS Macaca mulatta rhadinovirus 26-95.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=119193;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MACACA MULATTA RHADINOVIRUS ISOLATE 26-95;
RX MEDLINE=20173730; PubMed=10708456;
RA Alexander L., Denekamp L., Knapp A., Auerbach M.R., Damanian B.,
RT Desrosiers R.C.;
RT "The primary sequence of rhesus monkey rhadinovirus isolate 26-95;
RT sequence similarities to Kaposi's sarcoma-associated herpesvirus and
RT rhesus monkey rhadinovirus isolate 17577.";
RL J. Virol. 74:3388-3398(2000).
DR EMBL; AF210726; AAF59982.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR001230; Prenyltn.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
SQ SEQUENCE 395 AA; 43922 MW; BAC9C6F2E226AE06 CRC64;

Query Match      20.6%; Score 308.5; DB 12; Length 395;
Best Local Similarity 30.4%; Pred. No. 2e-21;
Matches 76; Conservative 40; Mismatches 113; Indels 21; Gaps 12;

QY 7 PPRNTEILTGSGWSQDTYPEGTQAIYKCRPG-YRSLGNVIMVC-RKGEWALNPLRKQK 64
DB 30 PFDFFMVKTN-QNENAVGTGRVELICRPGYKIQAVYVECLNGTWT--TPNAECRR 86

QY 65 RCPGHPGDTPTGFTLT--GGNVFEYGVKAVYTCNKGQYLLGEINREC---DTD--GWT 117
DB 87 KRCSNPEDILNGEVIITSDNAFKGNSITYKCNTRYLLG-ATVTRCLLYKYSNLVDWQ 145

QY 118 NDIPICEVVKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRVCNCGYKIEGDEEMHCS 176
DB 146 PAAPTCEIEKCKQPDIEGKYY-----PVQEFVNYLETITFTCNKDFSLIGNTTTTCWT 200

Query Match      20.6%; Score 308.5; DB 6; Length 522;
Best Local Similarity 29.8%; Pred. No. 2.9e-21;
Matches 84; Conservative 50; Mismatches 111; Indels 37; Gaps 17;

QY 3 CN---ELPPRRNTEILTGSGWSQDTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWALNPL 59
DB 48 CNAEQLPFPARTNLTDS-----EPVGYLYKECLPGYHGKPFSTICLKNSVMTSAKD- 102

QY 60 RKQKRCPCGHPGDTPTGFTLTGCVNFEYGVKAVYTCNKGQYLLGEINRECDDG---- 115
DB 103 -KCTRKSCRNPQDPVNGMYHVI--KDIQFGSQINYSCKNGYRLIGS-SSATCIISGNTVI 158

QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYH-FEQAVRVCNCG-----YKIE 167
DB 159 WDNETPICEIIPCGLPPTI-ANGDFISTS----REYFPYGVVYTYRCNLGSRGKLFELV 213

QY 168 GDEEMHCS--DD--GFWSKEPKC-VEISCKSPDVINGS-PIS-QKIYKENERFQYKCNM 221
DB 214 GEPSTYCKSDQVGIWSPAPQCIPKNCMPNPNVGVLSVYNRSLFSLNVEYFRCQP 273

QY 222 GYEYSERGDVCTE-SGWRP-LPSCSECKSCDNPYPNGDYSP 261
DB 274 GFVMKGPVRVQCALNKNWEPPLPSCSRVCPPEILLGHEHP 315

RESULT 12
ID Q99254 PRELIMINARY; PRT; 679 AA.
AC Q99254;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
```

```
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE=90229754; PubMed=2139460;
RA Kurtz C.B., O'Toole E., Christensen S.M., Weis J.H.;
RT "The murine complement receptor gene family. IV. Alternative splicing
RT of Cr2 gene transcripts predicts two distinct gene products that share
RT homologous domains with both human CR2 and CR1."
RL J. Immunol. 144:3581-3591(1990).
RN [2]
RP SEQUENCE OF 21-367 FROM N.A.
RX MEDLINE=95105691; PubMed=7528766;
RA Kim Y.U., Kinoshita T., Molina H., Hourcade D., Seya T., Wagner L.M.,
RA Holers V.M.;
RT "Mouse complement regulatory protein Crry/p65 uses the specific
RT mechanisms of both human decay-accelerating factor and membrane
RT cofactor protein."
RL J. Exp. Med. 181:151-159(1995).
DR EMBL; U17128; AAA78271.1; -.
DR EMBL; U17123; AAA78271.1; JOINED.
DR EMBL; U17124; AAA78271.1; JOINED.
DR EMBL; U17125; AAA78271.1; JOINED.
DR EMBL; U17126; AAA78271.1; JOINED.
DR EMBL; U17127; AAA78271.1; JOINED.
DR EMBL; M36470; AAA37449.1; -.
DR HSSP; P10998; 1VVD.
DR MGD; MGI:88489; Cr2.
DR InterPro; IPR002396; Selectin.
DR Pfam; PF00084; sushi; 10.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 10.
FT NON_TER 679
SQ SEQUENCE 679 AA; 74916 MW; 52FC00FDCED20CDC CRC64;

Query Match 20.1%; Score 301; DB 11; Length 679;
Best Local Similarity 26.0%; Pred. No. 2.1e-20;
Matches 82; Conservative 51; Mismatches 94; Indels 88; Gaps 16;

QY 20 SDQT-YPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLRKQCRPCGHPGDTPTGTF 78
DB 37 SOKSEFAIGTTWEYKCRPGYFRKSFITCLTSKW--SDAQOQFCKRKPCMNPOEPLHGSV 94
QY 79 TLTGGNVFEYGVKAVYTCNEGVLGELINVRQC---DTDGNTNDIPICEVVKCLPVTAP 134
DB 95 HINTG--IEFGSTITYSCNQYRLIGD--SSATCIVSDNTVMWMDNMPLECSIPCESPPAI 151
QY 135 ENKGISSAMEPDREYHFGQAVRVCNSG-----YKIEGDEEMHCDD-----GFWSKPK 184
DB 152 SNGDFYSSSRD---SPFYGMVYTYCHTKGNREKFLDLVGKSIYCTSKDNQVGINNSPP 208
QY 185 PKCV-EISCKSPDVG-----SPIS-----204
DB 209 POCIPRVKCPMEIENGLVSEGFKISFLNDVIFCKSGFTMKSGRIAMCOPNSKWSPP 268
QY 205 -----QKIY-----KENERF-----QYKCNMGYYSERGDVCTESG-W-RPL 241
DB 269 LPTCEMGLCPPQNILHGDYNNKDKDFSVGVGKVSYTCNPGCYTLIGTNLVCTSLGTWSNTV 328
QY 242 PSCEKSCNDPIPN 256
DB 329 PTCVKSCSD--AIPN 341

RESULT 13
```

```
Q9ES77
ID Q9ES77 PRELIMINARY; PRT; 3567 AA.
AC Q9ES77;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POLYDOM PROTEIN PRECURSOR.
GN POLYDOM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C3H/HENSIC;
RX MEDLINE=20517255; PubMed=11062057;
RA Gilges D., Vinit M.-A., Callebaut I., Coulombel L., Cacheux V.,
RA Romeo P.-H., Vigon I.;
RT "Polydom : a secreted protein with pentraxin, complement control
RT protein, epidermal growth factor and von willebrand factor A
RT domains."
RL Biochem. J. 352:49-59(2000).
DR EMBL; AF206329; AAG32160.1; -.
DR HSSP; P00740; 1EDM.
DR MGD; MGI:1928849; Polydom.
DR InterPro; IPR000152; Asx_hydroxyl.
DR EMBL; U17123; AAA78271.1; JOINED.
DR EMBL; U17124; AAA78271.1; JOINED.
DR EMBL; U17125; AAA78271.1; JOINED.
DR EMBL; U17126; AAA78271.1; JOINED.
DR EMBL; U17127; AAA78271.1; JOINED.
DR EMBL; M36470; AAA37449.1; -.
DR HSSP; P10998; 1VVD.
DR MGD; MGI:88489; Cr2.
DR InterPro; IPR002396; Selectin.
DR Pfam; PF00084; sushi; 10.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 10.
FT NON_TER 679
SQ SEQUENCE 679 AA; 74916 MW; 52FC00FDCED20CDC CRC64;

Query Match 20.1%; Score 300.5; DB 11; Length 3567;
Best Local Similarity 29.1%; Pred. No. 1.8e-19;
Matches 73; Conservative 34; Mismatches 119; Indels 25; Gaps 9;

QY 17 GSWSDQTYPEGTQAIYKCRPGYRSLGNVIMVC-RKGEWVALNPLRKQCRPCGHPGDTPT 75
DB 1797 GHSSGIYTVGTAVTFCDEGHVGVSTITCLTCEWDLRP--SCEATSCGP---PV 1851
QY 76 GFTLTGGNVFEYGVKAVYTCNEGVLGELINVRQC---DTDGNTNDIPICEVVKCLPVTAP 135
DB 1852 PNGGVGDGSAFTYGVKVVYKCDGYTLSDGESACIASGWSHSSPVLCVKCSQPEDIN 1911
QY 136 NGKIVSSAMEPDREYHFGQAVRVCNSG---YKIEGDEEMHCSDGGSWKEKPKVCVEISKSP 195
DB 1912 NGKYILSLGT-----YLSIASYSCENGYSLOQPSLLETASGSDWRAPPSCOLVSCGEP 1965
QY 196 DVINGSPISQKIYKENERF-----QYKCNMGYYSERGDVCTESG-WRPL-PSCEKSC 249
DB 196 DVINGSPISQKIYKENERF-----QYKCNMGYYSERGDVCTESG-WRPL-PSCEKSC 249
```

Db 1966 PIVKDA-----VITGSNFTFGNTVATCKEGYTLAGPDTIVCOANGKWNSSNHQCLAVSC 2020

QY 250 DNPYPINGDYS 260

Db 2021 DEP--PNVDHA 2029

RESULT 14

Q9UQV2

ID Q9UQV2 PRELIMINARY; PRT; 559 AA.

AC Q9UQV2;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE CRI C3b/C4b RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89010527; PubMed=2971757;

RA Hourcade D., Miesner D.R., Atkinson J.P., Holers V.M.;

RT "Identification of an alternative polyadenylation site in the human

RT C3b/C4b receptor (complement receptor type 1) transcriptional unit and

RT prediction of a secreted form of complement receptor type 1.";

RL J. Exp. Med. 168:1255-1270(1988).

DR EMBL; X14362; CAA32541.1;

DR HSSP; P10998; 1VVD.

DR InterPro; IPR002396; Selectin.

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR InterPro; IPR000834; zn-carbopept.

DR Pfam; PF00084; sushi; 8.

DR PRINTS; PR00343; SELECTIN.

DR SMART; SM00332; CCP; 8.

DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.

KW Signal; Receptor.

FT NON_TER 1

FT SIGNAL <1 16 POTENTIAL.

FT CHAIN 17 559 POTENTIAL.

SQ SEQUENCE 559 AA: 61424 MW; DBFFE965CA179D75 CRC64;

Query Match 19.9%; Score 298.5; DB 4; Length 559;

Best Local Similarity 27.0%; Pred. No. 2.8e-20;

Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;

QY 3 CNE---LPPRRNTEILTGSWSQTYPEGTQAIYKCRPGYRSLGNVTWCRKGEWALNPL 59

Db 18 CNAPEWLPARTNLT-----DEFEPIGYLYNECRPGYSGRPFSLCLKNSVWTGAKD- 72

QY 60 RKCQRKPCGHPGDTFGFTLTGGNVFEYGVKAVYTCNEGYSQLLGEINRYREC-----DTDG 115

Db 73 -RCRRKSCRNPDPVNGMVHVTKG--IQFGSQIKYSCTKGRLIGS--SSATCLISGDTVI 128

QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEDREYHFCQAVRVCNSG-----YKIEG 168

Db 129 WDNETPICDIRPCGLPPTT-TNGDFISTNRE---NFHYGSVYVTRCNPNGSGGRKVFELVG 184

QY 169 DEEMHC--SDD--GEWSKEKPKC-VEISCKSPDVLNGSPIS-QKIYKENERFOYKCNWG 222

Db 185 EFSIVCTSDNDQGVGWSGAPACIIPNCKTTPNVENGILVSDNLSLSLNFVEFRCPQG 244

QY 223 YEYSEGDVAVCTE-SGWRP-LPSCDEE-----KSCDNP 252

Db 245 FVMKGPVRVKCALNKEWELPSCSRVCQPPDPDLHAERTQDKDNFSPGQEVFYSCEPG 304

QY 253 Y-----IPNGDYS 261

Db 305 YDLRGAASMRCTPQGDWSP 323

RESULT 15

Q16745

ID Q16745 PRELIMINARY; PRT; 2039 AA.

AC Q16745;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE COMPLEMENT RECEPTOR 1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94065175; PubMed=8245463;

RA Vik D.P., Wong W.W.;

RT "Structure of the gene for the F allele of complement receptor type 1

RT and sequence of the coding region unique to the S allele.";

RL J. Immunol. 151:6214-6224(1993).

RN [2]

RP SEQUENCE FROM N.A.

RX Vik D.P., Wong W.W.;

RL Submitted (JUN-1993) to the EMBL/GenBank/DBDJ databases.

DR EMBL; L17418; AAB60694.1;

DR EMBL; L17390; AAB60694.1; JOINED.

DR EMBL; L17399; AAB60694.1; JOINED.

DR EMBL; L17409; AAB60694.1; JOINED.

DR EMBL; L17419; AAB60694.1; JOINED.

DR EMBL; L17420; AAB60694.1; JOINED.

DR EMBL; L17421; AAB60694.1; JOINED.

DR EMBL; L17422; AAB60694.1; JOINED.

DR EMBL; L17423; AAB60694.1; JOINED.

DR EMBL; L17391; AAB60694.1; JOINED.

DR EMBL; L17392; AAB60694.1; JOINED.

DR EMBL; L17393; AAB60694.1; JOINED.

DR EMBL; L17394; AAB60694.1; JOINED.

DR EMBL; L17395; AAB60694.1; JOINED.

DR EMBL; L17396; AAB60694.1; JOINED.

DR EMBL; L17397; AAB60694.1; JOINED.

DR EMBL; L17398; AAB60694.1; JOINED.

DR EMBL; L17400; AAB60694.1; JOINED.

DR EMBL; L17401; AAB60694.1; JOINED.

DR EMBL; L17402; AAB60694.1; JOINED.

DR EMBL; L17403; AAB60694.1; JOINED.

DR EMBL; L17404; AAB60694.1; JOINED.

DR EMBL; L17405; AAB60694.1; JOINED.

DR EMBL; L17406; AAB60694.1; JOINED.

DR EMBL; L17407; AAB60694.1; JOINED.

DR EMBL; L17408; AAB60694.1; JOINED.

DR EMBL; L17410; AAB60694.1; JOINED.

DR EMBL; L17411; AAB60694.1; JOINED.

DR EMBL; L17412; AAB60694.1; JOINED.

DR EMBL; L17413; AAB60694.1; JOINED.

DR EMBL; L17414; AAB60694.1; JOINED.

DR EMBL; L17415; AAB60694.1; JOINED.

DR EMBL; L17416; AAB60694.1; JOINED.

DR EMBL; L17417; AAB60694.1; JOINED.

DR HSSP; P08603; 1HFI.

DR InterPro; IPR001424; SOD_CU_ZN

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR InterPro; IPR000834; zn-carbopept.

DR Pfam; PF00084; sushi; 30.

DR SMART; SM00032; CCP; 30.

DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.

DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.

KW Receptor.

SQ SEQUENCE 2039 AA: 223603 MW; B82FCB11C6B16635 CRC64;

Query Match 19.9%; Score 298.5; DB 4; Length 2039;

Best Local Similarity 27.0%; Pred. No. 1.4e-19;

Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;

QY 3 CNE---LPPRRNTEILTGSWSQTYPEGTQAIYKCRPGYRSLGNVTWCRKGEWALNPL 59

Db 43 CNAPEWLPFAEPTNLTP---DEFEFFIGTYLNYECRPGYSGRPFISICLKNSVWTGAKD- 97
QY 60 RCKOKRPGCHPGDTFGTFTLTGGNVFEYGVKAVYTCNEGYOLLGEINYREC----DTDG 115
Db 98 -RCRRKSCRNPDPVNGVHVING--IQFGSQIKYISCTRGYRLIGS-SSATCIIISGDTVI 153
QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYHFGQAVRVCNSG-----YKIEG 168
Db 154 WNETPICDRIPCGLPPTI-TNGDFISTNRE---NFHYGSVVTYRCNPGSGGRKVFELVG 209
QY 169 DEEMHC--SDD--GFWSKXKPC-VEISCKSPDVINGSPIS-QKIIYKENERFOYKCNMG 222
Db 210 EPSIYCTSNDDQVGINSQPAQCIIPNKTTPNVENGILVSDNRSLSLNEVVEFEFCQFG 269
QY 223 YEYSERGDAVCTE-SGWRP-LPSCER-----KSCDNP 252
Db 270 FVMKGPRRYKCOALNKWKEPELFCSCRVQCPDVLHAERTORDKNFSPQOEVFYSCFPG 329
QY 253 Y-----IPNGDYSP 261
Db 330 YDLRGAAASMRCTPQGDWSP 348

Search completed: August 30, 2002, 06:22:07
Job time: 341 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2002, 06:04:46 ; Search time 29.78 Seconds
(without alignments)
988.400 Million cell updates/sec

Title: US-09-316-163-10

Perfect score: 1497

Sequence: 1 EDCNELPPRRTEILTGWS.....EKSCDNYPNGDYSPLRIK 265

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*

- 1: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/hold-geneseq/geneq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1055	70.5	240	18	AAW39154
2	961	64.2	216	18	AAW39155
3	320	21.4	645	21	AAAB53125
4	312	20.8	263	22	AAAB48846
5	303	20.2	581	12	AAAR13490
6	301.5	20.1	263	20	AAAY29859
7	301.5	20.1	263	21	AAAB13014
8	301	20.1	496	20	AAAY55752
9	300.5	20.1	543	13	AAAR28557
10	298.5	19.9	543	13	AAAR28547
11	298.5	19.9	543	13	AAAR28567

12	298.5	19.9	778	19	AAW73147	Amino acid sequenc
13	298.5	19.9	1930	19	AAW45899	Human complement r
14	298.5	19.9	2039	20	AAV55751	Human C3b/C4b rece
15	298.5	19.9	2039	22	ABG00287	Novel human diagno
16	298.5	19.9	2044	22	ABB11782	Human CRI protein
17	298.5	19.9	2044	22	AAW39224	Human polypeptide
18	298.5	19.9	2044	22	AAW41010	Human polypeptide
19	298.5	19.9	2317	10	AAV2219	CRI protein. Homo
20	296.5	19.8	543	13	AAAR28560	CRI-4 (114S) analo
21	295.5	19.7	543	13	AAAR28570	CRI-4 (266-274 KLK
22	294.5	19.7	543	13	AAAR28550	CRI-4 (64K) analog
23	294.5	19.7	543	13	AAAR28553	CRI-4 (85R, 87N) a
24	294.5	19.7	543	13	AAAR28565	CRI-4 (121Q) analo
25	294.5	19.7	543	13	AAAR28568	CRI-4 (347T, 349Y)
26	294.5	19.7	543	13	AAAR28569	CRI-4 (369-376 STK
27	294.5	19.7	543	13	AAAR28571	CRI-4 (364-367 NAA
28	294.5	19.7	2039	12	AAAR1810	Human complement t
29	293.5	19.6	543	13	AAAR28545	CRI-4 (37Y) analog
30	293.5	19.6	543	13	AAAR28548	CRI-4 (57V, 59K) a
31	293	19.6	363	18	AAW12414	Porcine complement
32	293	19.6	363	20	AAV30918	MCP protein. Unid
33	292.5	19.5	515	22	AAW93953	Human polypeptide,
34	292.5	19.5	543	13	AAAR28555	CRI-4 (92T) analog
35	292.5	19.5	543	13	AAAR28558	CRI-4 (109N, 110A,
36	292.5	19.5	543	13	AAAR28563	CRI-4 (117P) analo
37	292.5	19.5	1497	22	AAW93954	Human polypeptide,
38	291.5	19.5	543	13	AAAR28549	CRI-4 (64K, 65T) a
39	291.5	19.5	543	13	AAAR28551	CRI-4 (65T) analog
40	291.5	19.5	543	13	AAAR28562	CRI-4 (116K) analo
41	291.5	19.5	543	13	AAAR28566	CRI-4 (318K, 319N)
42	290.5	19.4	254	15	AAAR47154	Sequence of solubl
43	290.5	19.4	254	15	AAAR47155	Sequence of solubl
44	289.5	19.3	263	10	AAV2003	Deduced sequence o
45	289.5	19.3	543	13	AAAR28556	CRI-4 (94H) analog

ALIGNMENTS

RESULT 1
AAW39154
ID AAW39154 standard; Protein; 240 AA.
XX
AC AAW39154;
XX
DT 27-APR-1998 (first entry)
XX
DE Human partial Complement factor H protein fragment 1.
XX
DE Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator.
XX
OS Homo sapiens.
XX
PN WO9738136-A1.
XX
PD 16-OCT-1997.
XX
PF 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
PR 09-APR-1996; 96US-0015083.
PR 09-APR-1996; 96US-0630048.
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
PI Enfield DL, Hass GM, Kinders RJ;
XX
DR WPI: 1997-512742/47.
DR N-PSDB; AAV02790.
XX
PT Treating or screening for cancer, e.g. renal or urogenital cancer -

PT by modulating or detecting tumour associated human complement Factor
PT H related antigen, or nucleic acid encoding it
PS
PS Example 6B; Fig 6B; 104pp; English.
XX
CC This partial protein sequence represents a region of the human
CC tumour-associated complement factor H (CFH). This sequence is used
CC in the identification of complement factor H related proteins and
CC antigens isolated from clone PRBB9FH410 (see AAW39155). The detection of
CC such proteins and a CFH antigens can be used in screening or for the
CC treatment of renal or urogenital cancer, e.g. bladder, cervical or
CC prostate cancer. Agents that may modulate this antigen could be used in
CC the manufacture of a medicament for the treatment of a tumour cell.
XX
SQ Sequence 240 AA;

Query Match 70.5%; Score 1055; DB 18; Length 240;
Best Local Similarity 100.0%; Pred. No. 1e-80;
Matches 188; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 FTLTGGNVFYGKAVYTCNEGKOLLGEINVRCDTGDGWTNDIPICEVVKCLPVTAPENG 137
|||||
Db 1 flltggnvfygkavycnecgylgelnrctdgtwncdipicevkvclpvtapeng 60
|||||

QY 138 KIVSSAMEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFWSKEPKVCVEISCKSPDV 197
|||||
Db 61 kivssamepdrehyhfgqavrfvcnsgyklegdeemhcsddgfwskepkvcveisckspdv 120
|||||

QY 198 INGSPTSOIKIYKENERFOYKCNMGVEYSERGDVCTESGWRPLPSCEEKSCDNPYIPNG 257
|||||
Eb 121 inspslsqkilykenerfkykcmngveysergdavctesgwrplpsceekscdnpyipng 180
|||||

QY 258 DYSPLRIK 265
|||||
Db 181 dysplrik 188

RESULT 2
AAW39155
ID AAW39155 standard; Protein; 216 AA.
XX
AC AAW39155;
XX
DT 27-APR-1998 (first entry)
XX
DE Clone PRBB9FH410 CFH related protein fragment.
XX
KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator.
XX
OS Synthetic.
XX
PN WO9738136-A1.
XX
PD 16-OCT-1997.
XX
PF 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
PR 09-APR-1996; 96US-0015083.
PR 09-APR-1996; 96US-0630048.
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
PI Enfield DL, Hass GM, Kinders RJ;
XX
DR WPI; 1997-512742/47.
DR N-PSDB; AAW02791.
XX
XX Treating or screening for cancer, e.g. renal or urogenital cancer -
PT by modulating or detecting tumour associated human complement Factor

PT H related antigen, or nucleic acid encoding it
XX
PS Example 6B; Fig 6B; 104pp; English.
XX
CC This partial protein is found in clone PRBB9FH410 and represents a
CC complement factor H related protein with homology to a region of the
CC human tumour-associated complement factor H (CFH). The detection of this
CC protein and a CFH antigen can be used in screening or for the treatment
CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
CC Agents that may modulate this antigen could be used in the manufacture of
CC a medicament for the treatment of a tumour cell.
XX
SQ Sequence 216 AA;

Query Match 64.2%; Score 961; DB 18; Length 216;
Best Local Similarity 99.4%; Pred. No. 7e-73;
Matches 170; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 95 TCNEGKOLLGEINVRCDTGDGWTNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQ 154
|||||
Db 1 tcnegyqligelnrctdgtwncdipicevkvclpvtapengkivssamepdreynf9q 60
|||||

QY 155 AVRFVCSNGYKIEGDEEMHCSDDGFWSKEPKVCVEISCKSPDVINGSPISQKIIYKENER 214
|||||
Db 61 avrfvcnsgyklegdeemhcsddgfwskepkvcveisckspdvingspsqkiiykener 120
|||||

QY 215 FOYKCNMGVEYSERGDVCTESGWRPLPSCEEKSCDNPYIPNGDYSPLRIK 265
|||||
Db 121 fgykcmngveysergdavctesgwrplpsceekscdnpypngdysplrik 171
|||||

RESULT 3
AAB53125
ID AAB53125 standard; Protein; 645 AA.
XX
AC AAB53125;
XX
DT 28-FEB-2001 (first entry)
XX
DE Macaca mulatta rhadinovirus 17577 RRV ORF4 protein SEQ ID NO:7.
XX
KW Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
KW genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
KW cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KW splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.
XX
OS Macaca mulatta rhadinovirus 17577.
XX
PN WO200028040-A2.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26260.
XX
PR 06-NOV-1998; 98US-0107507.
PR 20-NOV-1998; 98US-0109409.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Wong SW, Axthelm MK, Searles RP;
XX
DR WPI; 2000-376552/32.
XX
XX New rhesus rhadino virus for producing non-human primate model useful
PT for testing potential treatments and efficacy of the candidate vaccine
PT for conditions associated with RRV infection -
XX
PS Claim 5; Page 122-123; 141pp; English.
XX
XX The present invention describes a novel rhesus macaque rhadinovirus

CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
 CC RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
 CC encoded by the genome sequence. The present invention also specifically
 CC claims the individual open reading frame (ORF) nucleotide sequences from
 CC the genome which encode the individual proteins, but these sequences are
 CC not given. A non-human animal infected with RRV can be used for testing
 CC the efficacy of drug in the treatment of condition associated with
 CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative
 CC disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
 CC hypergammaglobulinemia or autoimmune haemolytic anaemia, by
 CC administering the drug to an immuno-compromised non-human primate
 CC preferably Rhesus macaque monkey obtained by as a result of infection
 CC by Simian Immunodeficiency Virus (SIV). RRV is useful for producing
 CC non-human primate model for testing potential treatments for conditions
 CC associated with RRV infection. It is also useful for testing the
 CC efficacy of the candidate vaccine against RRV infection or conditions
 CC associated with its infection by administering the vaccine to the
 CC subject capable of infection with RRV, inoculating the subject with RRV
 CC and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205
 CC to AAB53213 represent sequence used in the exemplification of the
 CC present invention..
 XX

SQ Sequence 645 AA;

Query Match 21.4%; Score 320; DB 21; Length 645;
 Best Local Similarity 28.9%; Pred. No. 1.6e-18;
 Matches 82; Conservative 40; Mismatches 132; Indels 30; Gaps 12;
 QY 1 EDCNELPPRRNTEILTGSWSDQTYPEGTOAIYKRCRPGYSLGNVIMV--CRKGSEVALNP 58
 Db 21 enck--pphteyrvksntekdlysvgetaelicrpyvntkiilteclqngtw--stp 76
 QY 59 LKKQKRCPCGHPGDPFTFTLTGG--NVFEYGVKAVYTCNEGYQLLGEINRYREC---DTD 114
 Db 77 nfpcdkrcreptadllingavhhggdnalkfgsnisecnegydligs-nvrfcilqdt 135
 QY 115 --GWTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRFVCSGYKIEGDEE 171
 Db 136 nvnwdsnpevcclqckpavvehgdyi-----pqgdvnyngdaiftkcslytlvgst 190
 QY 172 MHCSDGFWSEKPKCKVEISCKSPDVING-SPTISQKIIYKENERFOYKCNMGYEYSESGD 230
 Db 191 lvctsnkksnsfptclmivcespqldngyidiglsrrynhqsitvksdgylnivgpet 250
 QY 231 AVCTESGW-RPLPSC-----EKKSCDNPYIPNGDYSPLRI 264
 Db 251 lictnttwypipkcvlvtmnpstmpetpmpetpdyqklnl 294

RESULT 4
 AAB48846
 ID AAB48846 standard; Protein; 263 AA.
 XX
 AC AAB48846;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Cowpox virus inflammation modulatory protein (IMP).
 XX
 KW Inflammation modulatory protein; IMP; cowpox virus;
 KW viral complement control protein; DAF homologue;
 KW complement activation inhibition; xenograft rejection.
 XX
 OS Cowpox virus.

XX WO200071147-A1.
 XX
 XX 30-NOV-2000.
 PD
 XX 24-MAY-2000; 2000WO-US14203.
 PF
 XX 25-MAY-1999; 99US-0136134.
 PR

XX
 PA (KING-) KING FAISAL SPECIALIST HOSPITAL & RES CE.
 PA (UYLO-) UNIV LOUISVILLE RES FOUND INC.
 XX
 PI Kotwal GJ, Al-Mohanna F, Parhar R;
 XX
 DR WPI; 2001-049865/06.
 DR N-PSDB; AAC87530.

Use of viral inflammation modulatory protein for blocking xenograft rejection in a patient

Claim 1; Page 26; 7lpp; English.

XX The invention relates to a novel method of preventing or treating
 CC xenograft rejection in a patient. The method comprises administering an
 CC inflammation modulatory protein (IMP) from cowpox virus, which acts as
 CC an inhibitor of complement activation. The invention also relates to
 CC a pharmaceutical composition comprising the IMP, suitable for the
 CC prevention or treatment of xenograft rejection. IMP is structurally
 CC related to the family of complement control proteins that includes DAF,
 CC IMP can block complement activation at a very early stage by binding to
 CC C3b and C4b, by inhibiting C3 convertase formation by either the
 CC classical or alternative pathway, and by accelerating the decay of the
 CC C3 convertase. Unlike other complement inhibitors, IMP retains most of
 CC the essential domains for complement control due to evolutionary
 CC pressure, and unlike sCRI, cleaves C3 at only the first cleavage site in
 CC the presence of factor 1. In addition, the viral protein is very small
 CC and is not recognised as a foreign protein because of its close
 CC structural similarity to its human homologues. The method and composition
 CC of the invention are useful for treating and preventing xenograft
 CC rejection in a patient in need of such treatment. The present sequence
 CC represents cowpox virus IMP.
 XX

SQ Sequence 263 AA;

Query Match 20.8%; Score 312; DB 22; Length 263;
 Best Local Similarity 33.1%; Pred. No. 2.5e-18;
 Matches 85; Conservative 31; Mismatches 111; Indels 30; Gaps 14;
 QY 3 CNELPPR-----RNTEILTGSWSDQTYPEGTOAIYKRCRPGYR--SLGNVIMVCRKGEWY 54
 Db 20 cctiparpinmkkfns---vgtdananyngidgtieylclpgyrkqkmgpiayctgtgwt 76
 QY 55 ALNPLRKCOKRCPGHPGDPFTFTLTGGNVFEYGVKAVYTCNEGYQLLGEI-NYRECDT 113
 Db 77 lfn---qcikrkcpspridngqldlg9---vefgssityscnsgyqligcsksycelgy 130
 QY 114 DG---WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVCSGYKIEGDE 170
 Db 131 tgsmvwnpeapicesvkcpssvntngr--hngyed--fytdgsvvtyscnsgysilgns 186
 QY 171 EMHCSDDGFWSEKPKCKVEISCKSPDVINGSPIQ-KIIYKENERFOYKCNMGYEYSESG 229
 Db 187 givcs-ggews-dpctcqvkcphptisngylssgfkrsyshndvdfkcrhyklsqss 244
 QY 230 DAVCTESG-WRP-LPSC 244
 Db 245 sstcspgntwqpelpkc 261

RESULT 5
 AAR13490
 ID AAR13490 standard; Protein; 581 AA.
 XX
 AC AAR13490;
 XX
 DT 30-OCT-1991 (first entry)
 XX
 DE Human C4 binding protein.
 XX

KW C4bp; monomer; complement protein; pJOD.C4bp.3; SCR;
 XX short consensus repeat.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..32

FT /label= signal_peptide

FT Protein 33..581

FT /label= C4bp

FT Region 33..93

FT /label= SCR8

FT Region 94..155

FT /label= SCR7

FT Region 156..219

FT /label= SCR6

FT Region 220..279

FT /label= SCR5

FT Region 280..345

FT /label= SCR4

FT Region 346..406

FT /label= SCR3

FT Region 407..464

FT /label= SCR2

FT Region 465..523

FT /label= SCR1

FT Domain 524..581

FT /label= C4bp_core

FT /note= "responsible for multimer assembly"

FT Disulfide-bond 34..80

FT /note= "intradomain"

FT Disulfide-bond 65..92

FT /note= "intradomain"

FT W09111461-A.

PN

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Best Local Similarity 30.7%; Pred. No. 3.9e-17;
 Matches 80; Conservative 42; Mismatches 101; Indels 38; Gaps 16;

QY 20 SDQTYEGTQAIYKCRPGY-RSLGNVIMVCRK-GEWALNPLRKCQKRPCHGDTFFGT 77

Db 51 tetrftgttlktyclpgyvrshstgtltcnsdgyw-yntf--cilykrcrhpg----- 102

QY 78 FTLTGGNV-----FEYGVKAVYTCNEGYQLLEINRECDT-----GWTNDIPICEVVKC 128

Db 103 --lrngqveiktdlsfgsqiefscgffligattsr-cevqrgvgywhlpqceivkc 159

QY 129 LPVTAPENGKIVSSAMEPDREYHFGQAVRVCNSGVKIEGDEMHCSDD----GFWSEK 184

Db 160 kppddlrngrhsge----enfaygfvstyscdprfslghaslsctvenetigvwrpsp 215

QY 185 PKCVEISCKSPDVINGSPISQ-KIIYKENERFOYKCNMGVEYSERGDAV--C-TESGWRP 240

Db 216 ptcekitcrkpdvshgemvsgfplynykdtlvfkqkgfvl--rgssvvhcdadskwp 273

QY 241 L-PSCEEKSCDN-PYIPNGDY 259

Db 274 sppacepncinlpdiphaw 294

RESULT 6

AAV29859

ID AAV29859 standard; Protein; 263 AA.

XX

AC AAV29859;

XX

DT 16-NOV-1999 (first entry)

XX

DE Vaccinia complement control protein sequence.

XX

KW Vaccinia virus; smallpox inhibitor of complement enzyme; SPICE;

KW fusion protein; hyperacute rejection; xenograft; inflammation;

KW post-ischaemic reperfusion injury; malignancy; autoimmune disease;

KW immune system disorder; neurodegeneration; infection; gene therapy;

KW blood additive; extracorporeal circulation system.

XX

OS Vaccinia virus.

OS Synthetic.

XX

PN W09944625-A1.

XX

PD 10-SEP-1999.

XX

PF 02-MAR-1999; 99WO-US04635.

XX

PR 03-MAR-1998; 98US-0076821.

XX

PA (UYJO) UNIV JOHNS HOPKINS.

XX

PA (UYPI-) UNIV PITTSBURGH.

XX

PI Rosengard AM, Ahearn JM;

XX

DR WPI; 1999-550981/46.

XX

DR N-PSDB; AA221091.

XX

PT New smallpox inhibitor of complement enzyme protein, used to treat

XX

PT complement-mediated disease, particularly hyperacute rejection -

XX

PS Claim 15; Fig 1; 88pp; English.

XX

XX The present invention describes the Vaccinia virus smallpox inhibitor of

CC complement enzymes (SPICE) protein. SPICE is an inhibitor of complement

CC activation, and so can be used to treat or prevent complement-mediated

CC disorders, especially hyperacute rejection, inflammation or post-

CC ischaemic reperfusion injury, malignancies, autoimmune diseases,

CC immune system disorders, neurodegeneration and infections. Hyperacute

CC rejection may also be prevented by treating the graft with SPICE before

CC transplanting it or by using a xenograft that has been transformed to

CC express SPICE from a gene therapy vector. SPICE is also useful as

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Query Match . 20.2%; Score 303; DB 12; Length 581;

Sequence 581 AA;

This sequence was deduced from human hepatocyte (Hep G2) cDNA

obtained following PCR amplification. The protein is a monomer

containing 8 SCRs. Each SCR forms a looped domain due to the

presence of two intradomain disulphide bonds (only the disulphide

bonds of SCR8 are labelled in the Features Table). Within each SCR,

the first cysteine residue bonds with the third and the second

cysteine residue bonds with the fourth. This secondary structure is

responsible for the conformational flexibility of the C4bp monomer.

The invention covers fusion proteins in which the monomer sequence,

or a specified subfragment of it, e.g. having 5, 4, 3 or 1 SCR(s)

is fused to the C-terminal of a protein such as a viral receptor,

cell ligand, a bacterial, viral or parasitic immunogen, enzyme,

cytokine, toxin, etc. See also AAQ13243-51.

Example 1; Fig 1; 105pp; English.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX


```

Db 289 ydlrgaasmrctpggdwspaaptcevkscdd 319
      Matches 88; Conservative 46; Mismatches 102; Indels 95; Gaps 18;

RESULT 10
AAR28547
ID AAR28547 standard; peptide; 543 AA.
XX
AC AAR28547;
XX
DT 19-MAR-1993 (first entry)
XX
DE CRI-4 (52S, 53S, 54P) analogue.
XX
KW short consensus repeat; regulator of complement activation;
KW C3b binding; C4b binding; human complement type 1 receptor.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1..60
FT /label= SCR-1
FT Region 61..122
FT /label= SCR-2
FT Region 451..510
FT /label= SCR-8
FT Region 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT Misc-difference 52
FT /note= "Thr substituted by Ser (SCR-8)"
FT Misc-difference 53
FT /note= "Gly substituted by Ser (SCR-8)"
FT Misc-difference 54
FT /note= "Ala substituted by Pro (SCR-8)"
XX
XX EP512733-A.
XX
PN 11-NOV-1992.
PD
XX
XX 28-APR-1992; 92EP-0303826.
XX
XX 03-MAY-1991; 91US-0695514.
XX
XX (UNIW ) UNIV WASHINGTON.
XX
XX Atkinson JP, Hourcade D, Krych M;
XX
XX WPI; 1992-375009/46.
XX
XX Complement activity regulator protein analogues - useful for
XX treating auto-immune diseases, to suppress transplant rejection,
XX for diagnosis etc.
XX
XX Claim 11; Fig 2 and R11810; 23pp; English.
XX
XX The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
XX 168:1255-1270. It encodes the first 8 and a half amino terminal
XX SCRs of CRI. The invention concerns analogues of "regulator of
XX complement activation" proteins or truncated, hybrid or recombinant
XX forms of them. CRI-4 is a preferred truncated form and a number of
XX specified substitution variants of it are claimed in which certain
XX positions in SCR-1 which have been identified as important for the
XX degree of C3b- and C4b-binding are substituted by amino acids from
XX the corresponding positions in SCR-8. The specification does not
XX contain the CRI-4 sequence; the sequence given here was constructed
XX from the full-length CRI amino acid sequence having GENESEQ
XX accession number AAR11810 and descriptions in the disclosure.
XX
XX Sequence 543 AA;

Query Match 19.9%; Score 298.5; DB 13; Length 543;
Best Local Similarity 26.6%; Pred. No. 8.5e-17;

Matches 88; Conservative 46; Mismatches 102; Indels 95; Gaps 18;
QY 3 CNE---LPPRRNTEILTGSMSDQTYPEGTQAIYKRCYRSLGNVIVCRKGEVVALNPL 59
Db 2 cnapwlpfarptnit-----defefpigtynyecrpysgrpfslclksv--sspk 55
QY 60 RKOKRRCGHPGDPFGTFTLTGNGVFEYGVKAVYTCNEGQQLLEINRYEC-----DTDG 115
Db 56 drcrkscrnpdpvngmvhvkq--lqfgsqikysctkgryllgs-ssatciliagdtv1 112
QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDRYHFGQAVRVCNSG-----YKIEG 168
Db 113 wdnetspicdrpcglpbt1-tngdfistnre---nfhygsvvtvrcnpgsggrkvfelvg 168
QY 169 DEEMHC--SD--GFWSKEKPKC-VEISCKSPDVINGSPIS-OKIIVKEN-----212
Db 169 epsiycstnddqvqgswgppaqcclpnkctppnvengllvsdnrlsflsnevefrqpv 228
QY 213 -----ERFQ-----YKCNMG 222
Db 229 fvmkgprvrkcqalnkwepepsepcsvcpqppdvhaertqrkdntfsggevfyscepg 288
QY 223 YEYSERGDAVCTESG-WRP-LPSCEEKSCDN 251
Db 289 ydlrgaasmrctpggdwspaaptcevkscdd 319

RESULT 11
AAR28567
ID AAR28567 standard; peptide; 543 AA.
XX
AC AAR28567;
XX
DT 19-MAR-1993 (first entry)
XX
DE CRI-4 (318-321 RNP) analogue.
XX
KW short consensus repeat; regulator of complement activation;
KW C3b binding; C4b binding; human complement type 1 receptor.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1..60
FT /label= SCR-1
FT Region 61..122
FT /label= SCR-2
FT Region 451..510
FT /label= SCR-8
FT Region 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT Misc-difference 318
FT /note= "Asp substituted by Arg"
FT Misc-difference 319
FT /note= "Asp substituted by Asn"
FT Misc-difference 320
FT /note= "Phe substituted by Pro"
FT Misc-difference 321
FT /note= "Met substituted by Pro"
XX
XX EP512733-A.
XX
PN 11-NOV-1992.
PD
XX
XX 28-APR-1992; 92EP-0303826.
XX
XX 03-MAY-1991; 91US-0695514.
XX
XX (UNIW ) UNIV WASHINGTON.
XX
XX Atkinson JP, Hourcade D, Krych M;
XX
XX

```


Query Match	19.98;	Score 298.5;	DB 20;	Length 2039;
Best Local Similarity	27.08;	Pred. No. 4.3e-16;		
Matches 86;	Conservative 49;	Mismatches 111;	Indels 73;	Gaps 18;

Qy	3	CNE---	LPPTRETELLTGSWSDQTYPECTQAIYKCRPGCYRSILGNVIMVCRKCEWVALNPL	59
Db	43	cnapewlpafapt	nt-----defefpigtynyecpysgrpsifclcnksvwtgaxd-	97
Qy	60	RKCKRRCPGHGPD	TPFTTLTGNVFYGVKAVYTCNCGYQLLGEINVREC----	115
Db	98	-rcrrkscrnppdvngm	vhvkg--lqfsgikysctkgyrlgs-ssatciisgdtvi	153
Qy	116	WTNDIPICEVVKC-	LPVTAPENGKIVSAMPDREYHFGQAVRVCNSG-----	168
Db	154	wdnetpicdrpcgl	pptti-tngdfistnre---nfhygsvvtvrcnpgsggrkvfelvg	209
Qy	169	DEEMHC--SDO-	-GFWSKEKPKC-VEISCKSPDVLINGSPI-SOKTIYKENERFOYKCNMG	222
Db	210	epslyctenddqvgi	wsgppqclpnkctppnvengllvdsnrslfslnevvefrcpg	269
Qy	223	YEYSERGDAVCTE-	SGWRP-LPSCEEE-----KSCNPP	252
Db	270	fvmkgprvrkca	lnkwepeipscsrvcqppdvlhaertardknfsgpgqevfyscpg	329
Qy	253	Y-----	IPNGDYSP	261
Db	330	ydlrgaasmrct	pqgdwsp	348

RESULT **15**

ABG00287

ID ABG00287 standard; Protein; 2039 AA.

XX

AC ABG00287;

XX

DT 13-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #278.

XX

KW Human; chromosome mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.

XX

PN WO200175067-A2.

XX

D 11-OCT-2001.

DD

XX	30 MAR 2001; 2001MC 0566031.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-APR-2000; 2000US-0649167.
XX	
XX	(HYSE-) HYSEQ INC.
XX	
XX	Drmanac RT, Liu C, Tang YT;
PI	
XX	
XX	WPI; 2001-639362/73.
DR	N-PSDB; AAS64474.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.
XX	
XX	
PS	Claim 20; SEQ ID No 30646; 103pp; English.

to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at www.wipo.int/pub/published_pct_sequences.

[illegible]

Search completed: August 30, 2002, 06:18:45
Job time: 839 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:02:15 ; Search time 39.15 Seconds
(without alignments)
807.494 Million cell updates/sec

Title: US-09-316-163-11

Perfect score: 1876

Sequence: 1 EDNCNLPPTRRNTEILTGSWS.....PDIKHGGLYHENMRPYPVPV 329

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1876	100.0	449	1 NBHUS	complement factor
2	1876	100.0	1231	1 NBHUS	complement factor
3	1261	67.2	1234	1 NBMSH	complement factor
4	903	48.1	669	2 S6551	factor H - bovine
5	486	25.9	1053	2 S46199	probable complemen
6	356	19.0	597	1 S53711	C4BP alpha chain p
7	353	18.8	597	1 NBHUC4	C4b-binding protei
8	353	18.8	676	2 A45900	complement C3b rec
9	350.5	18.7	2489	2 I73012	complement C3b/c4b
10	348	18.6	2014	2 I36936	complement recepto
11	343.5	18.3	482	2 A34924	complement C3b/C4b
12	337.5	18.0	560	2 T16833	hypothetical prote
13	335	17.9	360	2 T42921	complement control
14	323	17.2	497	2 J62054	complement regulat
15	320.5	17.1	610	1 I46001	C4b-binding protei
16	317	16.9	452	2 A35068	complement factor
17	317	16.9	558	2 S57953	C4BP protein alpha
18	311.5	16.6	469	1 NBMS064	C4b-binding protei
19	306	16.3	303	2 H35068	apolipoprotein H-r
20	304	16.2	440	2 A43519	complement recepto
21	303.5	16.2	1025	1 A43526	complement C3d/Eps
22	303	16.2	661	1 KFHUI3	coagulation factor
23	301.5	16.1	263	1 WMV2SP	apolipoprotein H h
24	294	15.7	668	2 A46013	coagulation factor
25	288	15.4	1091	1 P40009	complement C3d/Eps
26	286	15.2	830	2 A30359	P-selectin precurs
27	285	15.2	579	2 A56740	sperm-egg recognit
28	284.5	15.2	263	1 C36838	complement control
29	284.5	15.2	263	2 T28450	hypothetical prote

RESULT 1

NBHUS

complement factor H precursor, short splice form [validated] - human
N:Alternate names: Complement factor H-related protein; complement protein H
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 23-Feb-1996 #text_change 08-Dec-2000
C:Accession: S03013; B60238; A27877; A61103; A26505; S10479
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A:Title: The complete amino acid sequence of human complement factor H.
A:Reference number: S00254; MUID:88134059
A:Accession: S03013
A:Molecule type: mRNA
A:Residues: 1-449 <RIP>
A:Cross-references: EMBL:X07523; EMBL:Y00716; NID:g32492; PIDN:CAA30403.1; PID:g75807
A:Note: Part of this sequence, including the amino end of the mature protein was conf
A:Note: 402-Tyr was also found
R:Estaller, C.; Schwaible, W.; Dierich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A:Title: Human complement factor H: two factor H proteins are derived from alternativ
A:Reference number: A60238; MUID:91184292
A:Accession: B60238
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-33;434-449 <EST>
A:Note: only portions of this 1.8 kilobase mRNA were sequenced
R:Schulz, T.F.; Schwaible, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.
Eur. J. Immunol. 16, 1351-1355, 1986
A:Title: Human complement factor H: Isolation of cDNA clones and partial cDNA sequenc
A:Reference number: A27877; MUID:87054207
A:Accession: A27877
A:Molecule type: mRNA
A:Residues: 1-55-401, 'V', 403-449 <SCH>
A:Cross-references: GB:X04697; NID:g31991; PIDN:CAB41739.1; PID:g4725976
A:Note: an additional nucleotide present within the codon for Glu-310 was thought to
R:Schwaible, W.; Zwierner, J.; Schulz, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.
Eur. J. Immunol. 17, 1485-1489, 1987
A:Title: Human complement factor H: expression of an additional truncated gene produc
A:Reference number: A61103; MUID:88055295
A:Accession: A61103
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 27-76 <SC2>
A:Note: This is a partial sequence of an alternatively spliced 1.8 kilobase mRNA that
R:Sim, R.B.; Discipio, R.G.
Biochem. J. 205, 285-293, 1982
A:Title: Purification and structural studies on the complement-system control protein
A:Reference number: A26505; MUID:83048213
A:Accession: A26505
A:Molecule type: protein
A:Residues: 19-20,'Q',22-29,'V',31-33,'Q',35 <SIM>
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P

ALIGNMENTS

30	283.5	15.1	263	2	B72152
31	279.5	14.9	345	1	NBMS
32	279	14.9	610	2	A35046
33	276.5	14.7	612	2	B42755
34	274.5	14.6	349	2	G02913
35	274.5	14.6	369	2	I57998
36	274.5	14.6	768	2	A42755
37	274	14.6	377	2	I54479
38	274	14.6	384	2	S01896
39	274	14.6	768	2	I53821
40	270.5	14.4	362	2	JC5194
41	270.5	14.4	369	2	JC5138
42	269.5	14.4	360	1	WMBE2E
43	267.5	14.3	302	1	WMBE1E
44	267.5	14.3	345	1	NBHU
45	266	14.2	482	2	JC5092

B18L protein - var
apolipoprotein H p
E-selectin precurs
E-selectin precurs
Sperm CD46 - human
membrane cofactor
P-selectin precurs
membrane cofactor
membrane cofactor
P-selectin - rat
membrane cofactor
membrane cofactor
membrane-bound com
secretory compleme
apolipoprotein H p
E-selectin - pig

Biochemistry 31, 3626-3634, 1992
A:Title: Solution structure of the fifth repeat of factor H: A second example of the complement factor H precursor, long splice form [validated] - human
A:Reference number: A44551; MUID:92232649
A:Contents: annotation; NMR structure determination, residues 264-292
R:Kristensen, T.; Wetsel, R.A.; Tack, B.F.
J. Immunol. 136, 3407-3411, 1986
A:Title: Structural analysis of human complement protein H: homology with C4b binding protein
A:Reference number: S10479; MUID:86169701
A:Accession: S10479
A:Molecule type: mRNA
A:Residues: 226-401, 'Y', 403-449 <KRI>
A:Cross-references: GB:M12383; NID:g180472; PIDN:AAA52013.1; PID:g180473
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. Homologous proteins have also been found in liver, kidney, and placenta.
C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.
C:Genetics: <HF1>
A:Gene: GDB:HF1; HF
A:Map position: lq32-lq32
A:Cross-references: GDB:120041; OMIM:134370
C:Genetics: <HF2>
A:Gene: GDB:HF2; HF
A:Cross-references: GDB:129095
A:Map position: lq32-lq32
A:Note: the correspondence between the two loci and the sequences indicated is unclear; the alternative complement pathway
C:Function:
A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increases the alternative complement pathway
A:Pathway: complement alternate pathway
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-449/Product: complement factor H, short splice form #status experimental <MAT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357-401/Binding site: carbohydrate (Asn) #status absent

Query Match 100.0%; Score 1876; DB 1; Length 449;
Best Local Similarity 100.0%; Pred. No. 1.1e-126;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EDCNELPPRRNTILGSGNSDQTPPEGTQAIYKCRPGYSLGNVIMVCRKGWVALNPLR 60
DB 19 EDCNELPPRRNTILGSGNSDQTPPEGTQAIYKCRPGYSLGNVIMVCRKGWVALNPLR 78
QY 61 KCKRCPGHPDTPFTFTLTGNGVFEYGVKAVYTCNEGQYLLGEINRECDTGDGTNDI 120
DB 79 KCKRCPGHPDTPFTFTLTGNGVFEYGVKAVYTCNEGQYLLGEINRECDTGDGTNDI 138
QY 121 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVCSNGYKLEGDEMHCSDDGFW 180
DB 139 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVCSNGYKLEGDEMHCSDDGFW 198
QY 181 SKEKPKCVBEISCKSPDVINGSPISQIIYKNERFQYKCMGYEYSGRDAVCTESGWRP 240
DB 199 SKEKPKCVBEISCKSPDVINGSPISQIIYKNERFQYKCMGYEYSGRDAVCTESGWRP 258
QY 241 LPSCEKSCDNPYPINGDYSPLRKIRTKTGDEITYQCRNGFYPATRNTAKCTSTGWIPAP 300
DB 259 LPSCEKSCDNPYPINGDYSPLRKIRTKTGDEITYQCRNGFYPATRNTAKCTSTGWIPAP 318
QY 301 RCTLPCDYPDKHGGLYHENRRPFPV 329
DB 319 RCTLPCDYPDKHGGLYHENRRPFPV 347

RESULT 2

NBHUH

complement factor H precursor, long splice form [validated] - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence-revision 31-Dec-1993 #text-change 08-Dec-2000
A:Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A:Title: The complete amino acid sequence of human complement factor H.
A:Reference number: S00254; MUID:88134059
A:Accession: S00254
A:Molecule type: mRNA
A:Residues: 1-1231 <RIP>
A:Cross-references: EMBL:Y00716; NID:g31964; PIDN:CAA68704.1; PID:g31965
A:Note: 402-Tyr was also found
A:Note: parts of this sequence, including the amino and carboxyl ends of the mature
R:Estaller, C.; Schwaible, W.; Dierich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A:Title: Human complement factor H: two factor H proteins are derived from alternative
A:Reference number: A60238; MUID:91184292
A:Accession: A60238
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-56;1177-1231 <EST>
A:Note: only portions of this 4.3 kilobase mRNA were sequenced
R:Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.
Biosci. Rep. 7, 201-207, 1987
A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human comp
A:Reference number: A54726; MUID:88025472
A:Accession: A54726
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'DFRN', 579-1231 <DAY>
A:Cross-references: GB:M17517; NID:g180497; PIDN:AAA52016.1; PID:g180498
A:Note: parts of this sequence were determined by protein sequencing
R:Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.
Biosci. Rep. 6, 65-72, 1986
A:Title: Partial characterization of human complement factor H by protein and cDNA s
A:Reference number: A61565
A:Accession: A61565
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'METGRNHLNAKI', 1050-1057, 'T', 1059-1102 <RI2>
R:Sim, R.B.; Discipio, R.G.
Biochem. J. 205, 285-293, 1982
A:Title: Purification and structural studies on the complement-system control protei
A:Reference number: A26505; MUID:83048213
A:Accession: A26505
A:Molecule type: Protein
A:Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll,
Biochemistry 31, 3626-3634, 1992
A:Title: Solution structure of the fifth repeat of factor H: A second example of the
A:Reference number: A44551; MUID:92232649
A:Contents: annotation; NMR structure determination, residues 264-292
R:Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.
J. Mol. Biol. 219, 717-725, 1991
A:Title: Three-dimensional structure of a complement control protein module in solut
A:Reference number: A49224; MUID:91278097
A:Contents: annotation; NMR structure determination, residues 927-985
R:Estaller, C.; Kolstinen, V.; Schwaible, W.; Dierich, M.P.; Weiss, E.H.
J. Immunol. 146, 3190-3196, 1991
A:Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a n
A:Reference number: I56100; MUID:91201892
A:Accession: I72654
A:Status: translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1047-1231 <RES>
A:Cross-references: GB:M65294; NID:g183766; PIDN:AAA35948.1; PID:g183767
R:Carron, J.A.; Bates, R.C.; Smith, A.I.; Teto, T.; Arellano, A.; Gordon, D.L.; Bur
Biochim. Biophys. Acta 1289, 305-311, 1996
A:Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.
A:Reference number: S66298; MUID:96205365

A:Accession: S66298
A:Status: preliminary
A:Molecule type: protein
A:Residues: 411-419;574-578,580-582 <AR>
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. HC
C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.
C:Genetics: <HF1>
A:Gene: GDB:HF1; HF
A:Cross-references: GDB:120041; OMIM:134370
A:Map position: lq32-lq32
C:Genetics: <HF2>
A:Gene: GDB:HF2; HF
A:Cross-references: GDB:129095
A:Map position: lq32-lq32
A:Note: the correspondence between the two loci and the sequences indicated is unclear;
C:Function:
A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increased
he alternative complement pathway
A:Pathway: complement alternate pathway
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1229/Product: complement factor H #status experimental <MPT>
F:19-449/Product: complement factor H, short splice form #status experimental <MAT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:81-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:448-505/Domain: complement factor H repeat homology <FH08>
F:509-564/Domain: complement factor H repeat homology <FH09>
F:569-622/Domain: complement factor H repeat homology <FH10>
F:630-684/Domain: complement factor H repeat homology <FH11>
F:691-744/Domain: complement factor H repeat homology <FH12>
F:753-803/Domain: complement factor H repeat homology <FH13>
F:811-864/Domain: complement factor H repeat homology <FH14>
F:870-926/Domain: complement factor H repeat homology <FH15>
F:931-984/Domain: complement factor H repeat homology <FH16>
F:989-1043/Domain: complement factor H repeat homology <FH17>
F:1048-1163/Domain: complement factor H repeat homology <FH18>
F:1109-1163/Domain: complement factor H repeat homology <FH19>
F:1167-1228/Domain: complement factor H repeat homology <FH20>
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357-
1-803,811-853,839-864,870-915,901-926,931-973,959-984,989-1032,1018-1043,1048-1091,1077-
F:217/Binding site: carbohydrate (Asn) (covalent) #status absent
F:529,802,822,882,911/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:718,1029,1095/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 1876; DB 1; Length 1231;
Best Local Similarity 100.0%; Pred. No. 3.3e-126;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDCNELPRRNTTEILTGSWSQDTPEGTQALYKCRPGVRSIGNVIMVCKRGEWALNPLR 60
Db 19 EDCNELPRRNTTEILTGSWSQDTPEGTQALYKCRPGVRSIGNVIMVCKRGEWALNPLR 78
Qy 61 KCQRKPCGHPGDTTFFGTTTLTGGNVFEGYKAVYTCNBYQLLGEINRYECDDTGTWINDI 120
Db 79 KCQRKPCGHPGDTTFFGTTTLTGGNVFEGYKAVYTCNBYQLLGEINRYECDDTGTWINDI 138
Qy 121 PICEVVKCLPVTAPENGKIVSSAMEPREYHFGQAVRVCNSGYKIEGDEMHCSDDGFW 180
Db 139 PICEVVKCLPVTAPENGKIVSSAMEPREYHFGQAVRVCNSGYKIEGDEMHCSDDGFW 198
Qy 181 SKEPKKVEISCKSPDVINGSPISQKLIYKENERFOYKCNNGYYSERGDVAVCTESGWRP 240
Db 199 SKEPKKVEISCKSPDVINGSPISQKLIYKENERFOYKCNNGYYSERGDVAVCTESGWRP 258
Qy 241 LPSCEKSCDNPYPNGDYSPLRIKHTGTDEITYQCRNGFYPATRGNTAKTSTGTWIPAP 300

Db 259 LPSCEKSCDNPYPNGDYSPLRIKHTGTDEITYQCRNGFYPATRGNTAKTSTGTWIPAP 318
Qy 301 RCTLKPCDDYPDIKKHGGLYHENMRRPVPV 329
Db 319 RCTLKPCDDYPDIKKHGGLYHENMRRPVPV 347

RESULT 3
NBMSH
complement factor H precursor - mouse
N:Alternate names: protein beta-1-H
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 22-Jun-1999
C:Accession: A26154; I49711; I49728
R:Kristensen, T.; Tack, B.F.
Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986
A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in length
A:Reference number: A26154; MUID:86233353
A:Accession: A26154
A:Molecule type: mRNA
A:Residues: 1-1234 <KRI>
R:Natsume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.
J. Immunol. 144, 358-362, 1990
A:Title: Demonstration of an unusual allelic variation of mouse factor H by the compl
A:Reference number: I49711; MUID:90111033
A:Accession: I49711
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-18 <RES>
R:Munoz-Canoves, P.; Tack, B.F.; Vik, D.P.
Biochemistry 28, 9891-9897, 1989
A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma
A:Reference number: I49728; MUID:90148935
A:Accession: I49728
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-19 <RE2>
C:Cross-references: GB:J02891; NID:g193805; PIDN:AAA37795.1; PID:g553926
C:Comment: Two codominant alleles of factor H are present in mice.
C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prot
(C5 convertase) in the alternative complement pathway.

C:Genetics:
A:Map position: 1
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1234/Product: complement factor H #status predicted <MPT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:448-505/Domain: complement factor H repeat homology <FH08>
F:509-564/Domain: complement factor H repeat homology <FH09>
F:569-622/Domain: complement factor H repeat homology <FH10>
F:630-683/Domain: complement factor H repeat homology <FH11>
F:690-743/Domain: complement factor H repeat homology <FH12>
F:752-802/Domain: complement factor H repeat homology <FH13>
F:808-861/Domain: complement factor H repeat homology <FH14>
F:867-931/Domain: complement factor H repeat homology <FH15>
F:936-989/Domain: complement factor H repeat homology <FH16>
F:994-1048/Domain: complement factor H repeat homology <FH17>
F:1053-1107/Domain: complement factor H repeat homology <FH18>
F:1114-1168/Domain: complement factor H repeat homology <FH19>
F:1172-1233/Domain: complement factor H repeat homology <FH20>
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,
0-802,808-850,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,10

F:676,721,773,801,1030,1061,1225/Binding site: carbohydrate (Asn) (covalent) #status pre

Query Match 67.2%; Score 1261; DB 1; Length 1234;
Best Local Similarity 65.7%; Pred. No. 2.7e-82;
Matches 216; Conservative 37; Mismatches 76; Indels 0; Gaps 0;

QY 1 EDCNELPPRRNTEILTGSNDSTYPEGTAIKYKCRPGYSLGNVIMVCRKGEWALNPLR 60
DB 19 EDCKGPPPPRENSSEILGWSNQLYPEGTAIKYKCRPGYSLGNVIMVCRKGEWALNPLR 78

QY 61 KCQKRCCHGHPDTPGTFTLTCGNVFEYGVKAVYTCNEGQYLLGEINRECDTGTNDI 120
DB 79 ICRKKPCGHPDTPGTFTLTCGNVFEYGVKAVYTCNEGQYLLGEINRECDTGTNDI 138

QY 121 PICEVVKCLPVTAPENGKIVSSAMPEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFW 180
DB 139 PICEVVKCLPVTAPENGKIVSSAMPEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFW 198

QY 181 SKEKPKCVESCKSPDVINGSPISOKIIYKENERFYKCMGYEYSERGDVACTESGWRP 240
DB 199 SNEKPRCVIELTPPRVNGDGINVKPVYKENERVHYCKHGVVPKRGDVACTGSGWSS 258

QY 241 LPSCEKSCDNPIYINGDYSPLRIKHTGDEITYQCRNGFYFATRGNTAKCISTGTWIPAP 300
DB 259 QPFCBEKRCSPYILNGIYTPHRIIHRSDDEIRYECNMGFYFVGTGSTVSKCTPTGTWIPVP 318

QY 301 RCTLKPDPYDIKHHGLYHENRRPVPV 329
DB 319 RCTLKPDPYDIKHHGLYHENRRPVPV 347

RESULT 4
S65551
factor H - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S65551
R:Soames, C.J.; Day, A.J.; Slim, R.B.
Biochem. J. 315, 523-531, 1996
A:Title: Prediction from sequence comparisons of residues of factor H involved in the
A:Reference number: S65551; MUID:96202005
A:Accession: S65551
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-669 <SOA>
A:Cross-references: GB:X98697; NID:g1419423; PIDN:CAA67257.1; PID:g1419424
C:Superfamily: complement factor H; complement factor H repeat homology
F:55-114/Domain: complement factor H repeat homology <FHR1>
F:296-349/Domain: complement factor H repeat homology <FHR2>
F:355-412/Domain: complement factor H repeat homology <FHR3>
F:416-471/Domain: complement factor H repeat homology <FHR4>
F:476-530/Domain: complement factor H repeat homology <FHR5>
F:538-592/Domain: complement factor H repeat homology <FHR6>
F:599-651/Domain: complement factor H repeat homology <FHR7>

Query Match 48.1%; Score 903; DB 2; Length 669;
Best Local Similarity 60.9%; Pred. No. 5.2e-57;
Matches 154; Conservative 33; Mismatches 64; Indels 2; Gaps 1;

QY 76 GFTLTGNNVFEYGVKAVYTCNEGQYLLGEINRECDTGTNDIPICEVVKCLPVTAP 135
DB 3 GSPHLAENQFEGYKAVYTCNEGQYLLGEINRECDTGTNDIPICEVVKCLPVTAP 62

QY 136 NGKIVSSAMEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFWSKCKPKVEISCKSP 195
DB 63 NGKIVSSAMEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFWSKCKPKVEISCKSP 122

QY 196 DVINGSPISOKIIYKENERFYKCMGYEYSERGDVACTESGWRP LPSCEKSCDNPIY 255
DB 123 VILNGQAVLPKATYKQNERVYRCAAGFEYGGQGTGTCTKSGWTPAPTCTEITCDPRIP 182

QY 256 NGDYSPRLRIKHTGDEITYQCRNGFYFATRGNTAKCISTGTWIPAPRCTILKPCDYPDIKHG 315
DB 183 NGVYRPESLKYRGQDKITTYECKKGFPEIRGTATCTRDGWWVPVPRCAWKPCSYPIVIKHG 242

QY 316 GLYHENMRPYP 328
DB 243 RLYYS--YRGYFP 253

RESULT 5
S61699
probable complement regulatory plasma protein SBI - barred sand bass
C:Species: Paralabrax nebulifer
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 18-Jul-2001
C:Accession: S46199; S77894
R:Dahmen, A.; Kaideh, T.; Zipfel, P.F.; Gigli, I.
Biochem. J. 301, 391-397, 1994
A:Title: Cloning and characterization of a cDNA representing a putative complement-r
A:Reference number: S46199; MUID:94318039
A:Accession: S46199
A:Molecule type: mRNA
A:Residues: 1-1053 <DAH1>
A:Cross-references: EMBL:L21703; NID:g639894; PIDN:AAA92556.1; PID:g639895
A:Experimental source: liver
A:Accession: S77894
A:Molecule type: protein
A:Residues: 526-532, 'X', 534-537:809-817, 'X', 819-826 <DAH2>
C:Genetics:
A:Gene: SBI
C:Superfamily: complement factor H repeat homology
C:Keywords: glycoprotein
F:89-145/Domain: complement factor H repeat homology <FH01>
F:334-389/Domain: complement factor H repeat homology <FH02>
F:450-502/Domain: complement factor H repeat homology <FH03>
F:569-624/Domain: complement factor H repeat homology <FH04>
F:682-738/Domain: complement factor H repeat homology <FH05>
F:743-802/Domain: complement factor H repeat homology <FH06>
F:935-989/Domain: complement factor H repeat homology <FH07>
F:993-1052/Domain: complement factor H repeat homology <FH08>

Query Match 25.9%; Score 486; DB 2; Length 1053;
Best Local Similarity 29.7%; Pred. No. 5.1e-27;
Matches 114; Conservative 48; Mismatches 134; Indels 88; Gaps 13;

QY 21 DOTYPEGTAIKYKCRPGYSLGNVIMVCRKGEWALNPLRKCQKRCCHGHPDTPGTFTL 80
DB 47 EASYPGGRQVRVGCNNGYS--GFFKLCVYEGKWEIRG--AKQPRSCGHGPGDAQFADFL 102

QY 81 TCGNVFEYGVKAVYTCNEGQYLLGEINRECDTGTNDIPICEVVKCLPVTAPENGKIV 140
DB 103 AEGNDFVFGSKVYVTCQKGYQMVSRINRYRCVAEGMDGVVPCVCSQQCPLIHVDNNVQVI 162

QY 141 SSAMEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFWSKCKPKVEISCKSPDVIN 199
DB 163 GG-----PEATFGNVVRFSCRSRSEILDGSPELYCDERGDGSPVPCPKAKITCAIPPIEN 218

QY 200 GSPISOKIIYKENERFYKCMGYEYSERGDVACTESG-----WRPLPSCEKSC----- 249
DB 219 GNVPCAIREYKENDVLHYECDRAFKHIDR--PSTCIKQIKAEWSPPLCESIKRCLTMD 277

QY 250 -----DNPVIP-----NGOYS----- 260
DB 278 GTRYEPAYRNLFSPGETLKVICARTSWISTPOETSVYVTCQDNGEMSIPTCQEVRCNSR 337

QY 261 -PLRI-----KRTGDEITYQCRNGFYFATRGNT--AKCTSTGTWIPAPRCTILKPCDY 309
DB 338 REHVDSMDVRSWERYTDDNTRYCKRG--YKRTGGVTWATCGRNGWMPNPLCEVKTCSK 396

QY 310 PDIKHG-----GLYHENMRPYP 326
DB 397 ENIQDAVIGTDKQIYNLNQKAIY 420

RESULT 6

S53711
C4BP alpha chain precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S53711
R:de Frutos, P.G.; Dahlbaeck, B. 1995
Biochim. Biophys. Acta 1261, 285-289, 1995
A:Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence md
A:Reference number: S53711; MUID:95226458
A:Accession: S53711
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-597 <DEF>
A:Cross-references: EMBL:Z35490
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F:50-107/Domain: complement factor H repeat homology <FH1>
F:112-169/Domain: complement factor H repeat homology <FH2>
F:174-234/Domain: complement factor H repeat homology <FH3>
F:239-294/Domain: complement factor H repeat homology <FH4>
F:299-360/Domain: complement factor H repeat homology <FH5>
F:364-422/Domain: complement factor H repeat homology <FH6>
F:426-480/Domain: complement factor H repeat homology <FH7>
F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match 19.08; Score 356; DB 1; Length 597;
Best Local Similarity 30.5%; Pred. No. 5.3e-18;

Matches 106; Conservative 47; Mismatches 157; Indels 38; Gaps 20;

QY 2 DCNLEPPRRTEILTGSWSDTYPEGTQAIYKCRPGYRSLG-NVIMVCR-KGEWVALNPL 59
DB 49 DCGP-PHLLFASISLSENEYGTILKYTCRPGYRSLG-NVIMVCR-KGEWVALNPL 104
QY 60 RKCKRPGCHGPPFPGFTLTGCVNFEYGVKAVYTCNEGYQLLGEINRCCDPTD---G 115
DB 105 TFCVKKRCRPGDLPGQVEVK--TDFSGSQIEFSGEYILGST-SHCDLQEKVE 161
QY 116 WTNDIPICEVVKCLVPADPENGKIVSSAMEPDREYH-FGQAVRVCVSGYKIEDEEMHC 174
DB 162 WSDPLPKCEIVKCEPPNIIINGHNG---NEDIHTYGSSTVYSCNPFSLLEASISFC 217
QY 175 S---DDGFEWSKPKVEISKSPDVINGSPISO-KIIYKENERFOYKCNMGVEYSERG 229
DB 218 TVRKNKTGVNWSPPVCKEILCSPNPNVPHGKIIISGFGPIYKDSIMYTCIDGVL--RG 275
QY 230 DAV--C-TESGWRPLPS-CEEKSC-DNPIYIPNG-----DYSPLRIKHRTGDEITYQCRNG 279
DB 276 SSLIHCLELDKWNPPSPVCKESNCLGLPNVPHASQOQYQWSTREGVYVSGTELRYKCRPG 335
QY 280 FYP-ATRGNTAKC-TSGTWIPAPRCTLKPCDYPDIKIGGLYHENMRRP 325
DB 336 YRVADEPIITVCOEDLRWSPPFAGCEAICCPHPOLDNGAI-TEHSRNP 382

RESULT 7

NBUHCA
C4b-binding protein alpha chain precursor - human
N:Alternate names: C4BP; proline-rich protein
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1986 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: A33568; S02372; A90326; A24182; A93134; S29492; A31785; I52244; A03210
R:Matsuguchi, T.; Okamura, S.; Aso, T.; Sata, T.; Niho, Y.
Biochem. Biophys. Res. Commun. 165, 138-144, 1989
A:Title: Molecular cloning of the cDNA coding for proline-rich protein (PRP): identity of
A:Reference number: A33568; MUID:90073699
A:Accession: A33568
A:Molecule type: mRNA
A:Residues: 1-597 <NA1>
A:Cross-references: EMBL:Z35490; PIDN:AAA36507.1; PID:g190502
A:Note: The authors translated the codon GGA for residue 492 as Glu
R:Lintin, S.J.; Lewin, A.R.; Reid, K.B.M.

FEBS Lett. 232, 328-332, 1988
A:Title: Derivation of the sequence of the signal peptide in human C4b-binding protein
A:Reference number: S02372; MUID:88242821
A:Accession: S02372
A:Molecule type: mRNA
A:Residues: 17-81 <LI2>
A:Cross-references: EMBL:X07853
A:Note: although the sequence determined extends to residue 9 above, these authors as
R:Chung, L.P.; Bentley, D.R.; Reid, K.B.M.
Biochem. J. 230, 133-141, 1985
A:Title: Molecular cloning and characterization of the cDNA coding for C4b-binding pr
A:Reference number: A90326; MUID:86025405
A:Accession: A90326
A:Molecule type: mRNA
A:Residues: 80-597 <CH2>
A:Cross-references: GB:X02865; NID:g29564; PIDN:CAA26617.1; PID:g29565
R:Lintin, S.J.; Reid, K.B.M.
FEBS Lett. 204, 77-81, 1986
A:Title: Studies on the structure of the human C4b-binding protein gene.
A:Reference number: A24182; MUID:86301119
A:Accession: A24182
A:Molecule type: DNA
A:Residues: 203-288 <LI3>
A:Cross-references: EMBL:X04284; EMBL:X04296
R:Rodriguez de Cordoba, S.; Sanchez-Corrao, P.; Rey-Campos, J.
J. Exp. Med. 173, 1073-1082, 1991
A:Title: Structure of the gene coding for the alpha polypeptide chain of the human co
A:Reference number: A43023; MUID:91217619
A:Contents: annotation; exon-intron boundaries
R:Chung, L.P.; Gagnon, J.; Reid, K.B.M.
Mol. Immunol. 22, 427-435, 1985
A:Title: Amino acid sequence studies of human C4b-binding protein: N-terminal sequence
ced by cyanogen bromide treatment.
A:Reference number: A93134; MUID:85296001
A:Accession: A93134
A:Molecule type: protein
A:Residues: 49-81 <CH1>
A:Note: this paper reports amino-terminal sequences of the intact protein and of a N
R:Hessing, M.; Kanter, D.; Takeya, H.; van't Veer, C.; Hackeng, T.M.; Iwanaga, S.; B
FEBS Lett. 317, 228-232, 1993
A:Title: The region Ser(333)-Arg(356) of the alpha-chain of human C4b-binding protein
A:Reference number: S29492; MUID:93146164
A:Accession: S29492
A:Status: preliminary
A:Molecule type: protein
A:Residues: 381-404 <HES>
R:Suzuki, K.; Nishio, J.
J. Biol. Chem. 263, 17034-17039, 1988
A:Title: Binding site for vitamin K-dependent protein S on complement C4b-binding pro
A:Reference number: A31785; MUID:89034204
A:Accession: A31785
A:Molecule type: protein
A:Residues: 495-505, 'X', 507-510, 'X', 512-515 <SUZ>
A:Note: this peptide appears to bind protein S
R:Dahlback, B.; Smith, C.A.; Muller-Eberhard, H.J.
Proc. Natl. Acad. Sci. U.S.A. 80, 3461-3465, 1983
A:Title: Visualization of human C4b-binding protein and its complexes with vitamin K-
A:Reference number: A93950; MUID:83221615
A:Contents: annotation; electron microscopy; three-dimensional structure; ligand bind
R:Aso, T.; Okamura, S.; Matsuguchi, T.; Sakamoto, N.; Sata, T.; Niho, Y.
Biochem. Biophys. Res. Commun. 174, 222-227, 1991
A:Title: Genomic organization of the alpha chain of the human C4b-binding protein gen
A:Reference number: I52244; MUID:91113199
A:Accession: I52244
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-597 <ASO>
A:Cross-references: GB:M62486; NID:g190498; PIDN:AAA36506.1; PID:g190500
C:Comment: C4BP controls the classical pathway of complement activation. It binds as
the degradation of the C4bC2a complex (C3 convertase) by dissociating the complement
C:Comment: C4BP occurs in plasma in two forms, both of which bind complement fragment
isulfide bonds. The beta chain binds the vitamin K-dependent plasma protein S. Amino

actors V and VIII.
C:Comment: The molecule has a central body supporting seven tentacles (alpha chains), ea
C:Genetics:
A:Gene: GDB:C4BPA
A:Cross-references: GDB:120568; OMIM:120830
A:Map position: 1q32-1q32
A:Introns: 48/1; 110/1; 143/2; 172/1; 236/1; 297/1; 362/1; 425/1; 482/1; 540/3
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein; pla
F:1-48/Domain: signal sequence #status predicted <SIG>
F:49-59/Product: C4b-binding protein alpha chain #status predicted <MAT>
F:50-108/Domain: complement factor H repeat homology <FH1>
F:113-170/Domain: complement factor H repeat homology <FH2>
F:175-234/Domain: complement factor H repeat homology <FH3>
F:239-294/Domain: complement factor H repeat homology <FH4>
F:299-360/Domain: complement factor H repeat homology <FH5>
F:364-422/Domain: complement factor H repeat homology <FH6>
F:381-404/Region: complement C4b binding #status predicted
F:426-480/Domain: complement factor H repeat homology <FH7>
F:484-538/Domain: complement factor H repeat homology <FH8>
F:221,506,528/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 18.8%; Score 353; DB 1; Length 597;
Best Local Similarity 29.3%; Pred. No. 8.7e-18;
Matches 100; Conservative 50; Mismatches 141; Indels 50; Gaps 21;

Qy 20 SDQTYPEGTQAIYKRCPCY-BSLGNVIMVCRK-GEWVALNPLRKCKRPGCHGPDTPFGT 77
Db 67 TETRFKGTTLKYTLCLPGVSHSTQTUTLCSNGGEW-YNTF-CIYARCRHGE----- 118
Qy 78 FTLTGGNV-----FEYGVKAVYTCNEGYYQLLGEINRYREC-----GTNDIPICEVYK 128
Db 119 --LRNQVEIKTDLFSQIEFSCQIEFSCQIEFSCQIEFSCQIEFSCQIEFSCQIEFSCQ 175
Qy 129 LPVTAPENKIVSSAMEPDREYHFGQAVRFVNCNGYKTEGDEEMHCSD-----GFWSKEK 184
Db 176 KPPDPDIRNGRSGE-----ENFYAGFVSYTSCDPFSLGHASISCTVENETIGVWRPSP 231
Qy 185 PKCIVEISCKSPDVINGSPISO-KIYKENERFYKCNMGYSEYSEKDAV--C-TESGWRP 240
Db 232 PTCEKITCRKPDVSHGENVSGFGPIYNYKDTIVFKCKQGFVL--RGSSVIHCADSKNPP 289
Qy 241 L-PSCEEKSCDN-PYIPNGDYS--PLRIK---HRTGDEITYOCNRYGFYAT-RGNTAKC- 291
Db 290 SPPACEPNSCINLDPHASEWETPRPTKEDYVVGTVLVRCHPGYKPTTDEPTVICQ 349
Qy 292 TSTGWIPAPRCTLPKCDYDKHGLYHENMRP-----YF 327
Db 350 KNLRTYPOGCEALCCPEKLNNGEITQHRKSRPANHCYVF 390

RESULT 8
A45900
Complement C3b receptor type 2 long form precursor - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C:Accession: A45900; I48306
R:Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.
J. Immunol. 144, 3581-3591, 1990
A:Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2 gene
A:Reference number: A45900; MUID:90229754
A:Accession: A45900
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-676 <KUR>
A:Cross-references: GB:M36470
A:Experimental source: clone 31-1
R:Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers, V.
J. Exp. Med. 181, 151-159, 1995
A:Title: Mouse complement regulatory protein Cr2/p65 uses the specific mechanisms of bo
A:Reference number: I48306; MUID:95105691

A:Accession: I48306
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 21-367 <RES>
A:Cross-references: EMBL:U17128; NID:9595980; PIDN:AAA78271.1; PID:9595982
C:Genetics:
A:Gene: Cr2
A:Introns: 80/1; 113/2; 142/1; 275/1; 333/1
C:Superfamily: complement factor H repeat homology
F:22-78/Domain: complement factor H repeat homology <FH01>
F:83-140/Domain: complement factor H repeat homology <FH02>
F:145-211/Domain: complement factor H repeat homology <FH03>
F:217-272/Domain: complement factor H repeat homology <FH04>
F:276-331/Domain: complement factor H repeat homology <FH05>
F:336-394/Domain: complement factor H repeat homology <FH06>
F:399-458/Domain: complement factor H repeat homology <FH07>
F:467-523/Domain: complement factor H repeat homology <FH08>
F:531-587/Domain: complement factor H repeat homology <FH09>
F:592-648/Domain: complement factor H repeat homology <FH10>

Query Match 18.8%; Score 353; DB 2; Length 676;
Best Local Similarity 28.9%; Pred. No. 1e-17;
Matches 93; Conservative 61; Mismatches 136; Indels 32; Gaps 17;

Qy 20 SDQTYPEGTQAIYKRCPCYBSLGNVIMVCRKGEWVALNPLRKCKRPGCHGPDTPFGT 78
Db 37 SDKSEFALGTWYKACRCGYPRKSFIIITCLESKW--SDAQFCRKRKPCMNPOEPLGVS 94
Qy 79 TLTGNGVEYGVKAVYTCNEGYYQLLGEINRYREC-----DTGWTNDIPICEVYKLPVTAP 134
Db 95 HINTG--IEFGSTIYSCNQYRLIGD--SSATCIYSDNTVMYMDNDMPLESPCESPPAI 151
Qy 135 ENKIVSSAMEPDREYHFGQAVRFVNCNGYKTEGDEEMHCSD-----YKTEGDEEMHCSD 184
Db 152 SNGDFYSSRD---SFFYGVWVYTYCHTGKREKLEFLDVGESKIYCTSKDNQVGIWNSPP 208
Qy 185 PKCV-EISCKSPDVINGSPISO-KIYKENERFYKCNMGYSEYSEKDAV--TESGWR-P 240
Db 209 PQCIPIRVKCPMEIENGLVESGFKHSFFLNDTVIFKCKSGFTMKGSRIACQPNKWSPP 268
Qy 241 LPSCEEKSCDN-PYIPNGDYSPLRIKHRTGDEITYOCNRYGFYATRGNTAKCTSTG-WI-P 298
Db 269 LPTCFMGLCPQNLHGDYNNKDEFFSVGQKVSVCNFG-YTLIGTLNLECTSLGTSMTSNT 327
Qy 299 APRCTLKPCD-YPD-IKHGGLY 318
Db 328 VPTCEVKSCDAIPNHLHGRVF 349

RESULT 9
I73012
Complement C3b/C4b receptor, membrane-bound form precursor - human
N:Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CRI); su
N:Contains: complement C3b/C4b receptor, secreted form
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 24-Nov-1999 #text_change 21-Jul-2000
C:Accession: I73012; I56203; A47602; S03291; S03843; A28507; A24748; C24748
R:Vik, D.P.; Wong, W.W.
J. Immunol. 151, 6214-6224, 1993
A:Title: Structure of the gene for the F allele of complement receptor type 1 and se
A:Reference number: I56203; MUID:94065175
A:Accession: I73012
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-683, 'X', 685-1133, 'X', 1135-1471, 'X', 1473-2489 <VIK1>
A:Cross-references: GB:LI7418; NID:9306678; PIDN:AAB60695.1; PID:g451303
A:Accession: I56203
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-683, 'X', 685-894, 'A', 896-1000, 1451-1471, 'X', 1473-2489 <VIK2>
A:Cross-references: GB:LI7418; NID:9306678; PIDN:AAB60694.1; PID:9306680
R:Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.,

F: 1236-1291/DomaIn:	complement	H repeat	homology	<FH13>
F: 1236-1362/DomaIn:	complement	factor	homology	<FH14>
F: 1237-1489/DomaIn:	complement	factor	homology	<FH15>
F: 1245-1681/DomaIn:	complement	factor	homology	<FH16>
F: 1749-1835/DomaIn:	complement	factor	homology	<FH17>
F: 1823-1879/DomaIn:	complement	factor	homology	<FH18>
F: 1884-1950/DomaIn:	complement	factor	homology	<FH19>

23;

QY	184	KPRC	VEIS-CKSPD	VINGSPI-SQR	IYK	ENRFOYKCNMGEYSERGDVACTESG-WRP	240
		: :	: :	: :	: :	: :	
Db	1557	PPRC	ISTNCTA	PEVENATRV	PGNSFFSLTEIVFR	COPGFVWGSITVCQTGRWGP	1616
		: :	: :	: :	: :	: :	
QY	241	-LPS	EKSCDN	PYPNGDY	SPLRIKHR----	TGDEITYQCRNGFYPATRGNTAKACTSTG	295
				: :		: :	
Db	1617	KLPHCS	RVCOPPEILH	GHEHTP----	SHQDNFSGOEVFVSCEPG-YDLRGAASLHCRTPOG		1672
				: :		: :	
QY	296	-WIP-	APRCTLP	PCD--YPD	IKHGGLYHENMRPYFPV		329
				: :		: :	
Db	1673	DWSPE	APRCTLVKSCDD	FLGQLPHG-----	RVLFPL		1702
				: :		: :	

60 -----

A; Cross-references: EMBL:X14360

Db	73	LQLKCGPPPEIPFAVHDGSSFGSEYDLDAEGLSISKCLLNKRKNVAQWQFGPDLURCKARACP	133
Qy	69	HPGDTPTGTFTLTGGNVFYEYKAVVTCNEGYYOLLCEINRECDTDG-WTNDIPICEVVK	127
Db	133	DPGDIENG---LRGDTFEPHHVYKSCNPGFLLVGST-ROCSSNGEWTPANCKKATE	188
Qy	128	CLPVTAPENKIVSSAMEPDREYHFGQAVRFVCNCSYKIEGDEEHCHSCDDGFWSEKPKC	187
Db	189	CSRSPSLPHGKVVGSILT-----YQSVVYSCDHCYRLVGQVQRICLAELGTWGGNEPC	242
Qy	188	VEISCK-SPDVIENG-----SPI-----	203
Db	243	EEIRCSVLPTLPNGYIEGSETSFQAVAFRCLETMTHEGASKAKMEDQWSAPIRCLA	302
Qy	204	SOKIYKENERFOYK-----CNMGYEYSERGDAVCTESGWRPLPSCSEKS	248
Db	303	SCRVPHQNGKIKDKSEGLIASGKVIYECNKQHEANTDERLICSNSTWSHVPCVSLPS	362
Qy	249	CDN---PYIPNGDYSPLRIKHRTGDEITYOCRNNGFYFATRGNTAKCTSTGWI---PAPRPT	303
Db	363	CHNWPVRVPHARI--LFSKSSHGSIKAYECNNGYGHFNRNNQIIKCLYGTEWKDGPMPKCL	420
Qy	304	LKPCDYDPIKHGGL	317
Db	421	PSWCEHPDSKTYGTL	434

```

RESULT 13
T42921
complement control protein homolog ccph - ateline herpesvirus 3 (strain 73);
C:Species: ateline herpesvirus 3
A:Variety: strain 73
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42921
R:Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A:Description: Primary structure of the herpesvirus ateles genome.
A:Reference number: Z22274
A:Accession: T42921
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-360 <ALB>
A:Cross-references: EMBL:AF083424; PIDN:AAC95530.1
A:Experimental source: strain 73

```

Query Match 17.98; Score 335; DB 2; Length 360;
Best Local Similarity 33.4%; Pred. No. 9.8e-17;
Matches 103; Conservative 32; Mismatches 119; Indels 54; Gaps 20;

[illegible]

Db 291 CNOQNTTS 298

RESULT 14
JC2054

complement regulatory protein, S12 antigen precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 21-Jul-2000
C:Accession: JC2054; PC2027
R:Sakurada, C.; Seno, H.; Dohi, N.; Takizawa, H.; Nonaka, M.; Okada, H.
Biochem. Biophys. Res. Commun. 198, 819-826, 1994
A:Title: Molecular cloning of the rat complement regulatory protein, S12 antigen.
A:Reference number: JC2054; MUID:94161746
A:Accession: JC2054
A:Molecule type: mRNA
A:Residues: 1-497 <SAA>
A:Cross-references: GB:D42114; NID:g577641; PIDN:g987688
A:Accession: PC2027
A:Molecule type: protein
A:Residues: 39-51 <SA2>
C:Comment: This protein plays a critical role in protection against complement mediated
C:Superfamily: C3b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: glycoprotein; transmembrane protein
F:1-36/Domain: signal sequence #status predicted <SIG>
F:37-468/Product: complement regulatory protein, S12 antigen #status predicted <NAT>
F:38-94/Domain: complement factor H repeat homology <FH1>
F:99-156/Domain: complement factor H repeat homology <FH2>
F:161-227/Domain: complement factor H repeat homology <FH3>
F:233-288/Domain: complement factor H repeat homology <FH4>
F:294-352/Domain: complement factor H repeat homology <FH5>
F:357-413/Domain: complement factor H repeat homology <FH6>
F:421-444/Domain: transmembrane #status predicted <TMN>
F:247,331,346,450,482,483/Binding site: Carbohydrate (Asn) (covalent) #status predicted

```

Query Match      17.2%   Score 323;   DB 2;   Length 497;
Best Local Similarity 28.3%;   Pred. No. 9.9e-16;
Matches 95;   Conservative 55;   Mismatches 144;   Indels 42;   Gaps 21;

QY  21  DOTYPECTQAIYKCRPGYRSIGNVIMYCRKGEWALNPLRKCKQKRCGHPGDTPFGFTL 80
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   55  ESTFPVTSYKUECRPGYIKRQFSITCEVNSVMT--SPQDVCIRKOCETPLDPONGIVH 112
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  81  TGGNVFYGKAVYTCNEGYSQQLLEINRYBCDDT--GWTNDIPICEVVK-LPVTAPEN 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   113 --NTDIRFGSIYTCNEGYSRLIGSSAMCIISQSVANDAEAPICESICEIPPSIP-N 169
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  137 GKIVSSAMEPDRE-YHFGQAVRFVCSNG-----YKIEGDEMHCS----DDGFWSEKP 185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   170 GDFES----PNREDFHYGMVTYQCNTDARGKKLFNLVGEPSIHCTSIDGOVCWWSGPPP 225
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  186 KCVETIS-CKSPDVLINGSPIQ-KIIYKENERFOYKCNMGYEYSERGDVCTE-SGRHP-L 241
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   226 QCIELNCTPPHVENAVIVSKNLSLRLDMVFECODGFMFKGDSVYGRSLNRWEPQL 285
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  242 PSC-EERKSCDN--PYIPNGD-YSPLRIKHRTGEITYQCRNGFYPATRGNTAK-CTSTG- 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   286 PSCFKYKSCGAFLGELPNGHVFPQNL--QLGAKVTFVCNTGY--QLKGNSSSHCVLDGV 341
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  296 --W-IPAPRCTLKPCDYPDIKHGGLYHENRRPYF 327
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   342 ESTWNSVPVCEQVICKLPDMMSGFQKGLQMKDDY 377
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15
I46001
C4b-binding protein alpha chain - bovine
C:Species: Bos primigenius taurus (cattle)
C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: I46001; S43190
R:Hillarp, A.; Thern, A.; Dahlback, B.
J. Immunol. 153, 4190-4199, 1994
A:Title: Bovine C4b binding protein. Molecular cloning of the alpha- and beta-chains

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:12:04 ; Search time 21.93 Seconds*
(without alignments)
580.882 Million cell updates/sec

Title: US-09-316-163-11
Perfect score: 1876
Sequence: 1 EDCNELPPRRNTEILTGSWS.....PDIKHGGLYHENRRPFPV 329

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1876	100.0	1231	1	CFAH_HUMAN
2	1261	67.2	1234	1	CFAH_MOUSE
3	927.5	49.4	685	1	CFAH_BOVIN
4	353	18.8	597	1	C4BP_HUMAN
5	350.5	18.7	2039	1	CR1_HUMAN
6	320.5	17.1	610	1	C4BP_BOVIN
7	317	16.9	558	1	C4BP_RAT
8	311.5	16.6	489	1	C4BP_MOUSE
9	307	16.4	1033	1	CR2_HUMAN
10	303.5	16.2	1025	1	CR2_MOUSE
11	303	16.2	661	1	F13B_HUMAN
12	301.5	16.1	263	1	VCP_VACCV
13	294	15.7	668	1	F13B_MOUSE
14	286	15.2	830	1	LEM3_MOUSE
15	279.5	14.9	345	1	APDH_MOUSE
16	279	14.9	610	1	LEM2_HUMAN
17	276.5	14.7	612	1	LEM2_MOUSE
18	274.5	14.6	768	1	LEM3_MOUSE
19	274	14.6	377	1	MCP_HUMAN
20	274	14.6	768	1	LEM3_RAT
21	270.5	14.4	611	1	LEM2_CANFA
22	269.5	14.4	360	1	CCPH_HSVSA
23	267.5	14.3	345	1	APDH_HUMAN
24	266	14.2	484	1	LEM2_PIG
25	262.5	14.0	958	1	HIG_DROME
26	262	14.0	381	1	DAF_HUMAN
27	261.5	13.9	345	1	APDH_BOVIN
28	259	13.8	769	1	LEM3_SHEEP
29	252	13.4	340	1	DAF_PONPY
30	246.5	13.1	507	1	DAF_CAVPO
31	242.5	12.9	345	1	APDH_CANFA
32	241	12.8	549	1	LEM2_RAT
33	239.5	12.8	551	1	LEM2_RABIT

ALIGNMENTS

RESULT 1

ID	CFAH_HUMAN	STANDARD;	PRT;	1231 AA.
AC	P08603; Q14570; P78435; Q9NU86;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Complement factor H precursor (H factor 1).			
DE	HF1 OR HF OR CFH.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT Y-402.			
RC	TISSUE=Liver;			
RX	MEDLINE=88134059; PubMed=2963625;			
RA	Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;			
RT	*The complete amino acid sequence of human complement factor H.;			
RL	Biochem. J. 249:593-602(1988).			
RN	[2]			
RP	SEQUENCE OF 53-445 FROM N.A.			
RX	MEDLINE=87054207; PubMed=2946589;			
RA	Schulz T.F., Schwaible W., Stanley K.K., Weiss E., Dierich M.P.;			
RT	*Human complement factor H: isolation of cDNA clones and partial cDNA sequence of the 38-kDa tryptic fragment containing the binding site for C3b.;			
RL	Eur. J. Immunol. 16:1351-1355(1986).			
RN	[3]			
RP	SEQUENCE OF 226-449 FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=86169701; PubMed=2937845;			
RA	Kristensen T., Wetsel R.A., Tack B.F.;			
RT	*Structural analysis of human complement protein H: homology with C4b binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.;			
RL	J. Immunol. 136:3407-3411(1986).			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM 2).			
RA	Bird C.;			
RT	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	SEQUENCE OF 1047-1231 FROM N.A.			
RX	MEDLINE=91201892; PubMed=1826708;			
RA	Estaller C., Koistinen V., Schwaible W., Dierich M.P., Weiss E.H.;			
RT	*Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel member of the short consensus repeat family related to the carboxy terminal of the classical 150-kDa molecule.;			
RL	J. Immunol. 146:3190-3196(1991).			
RN	[6]			
RP	SEQUENCE OF 19-35.			
RX	MEDLINE=83048213; PubMed=6215918;			
RA	Sim R.B., Discipio R.G.;			
RT	*Purification and structural studies on the complement-system control protein beta 1H (Factor H).;			
RL	Biochem. J. 205:285-293(1982).			
RN	[7]			
RP	SEQUENCE OF 1-19 FROM N.A.			

Q61475 mus musculus
P98107 bos taurus
Q92496 homo sapien
Q02985 homo sapien
Q03591 homo sapien
Q61476 mus musculus
P42201 bos taurus
P25644 rattus norv
P18337 mus musculus
P14151 homo sapien
Q95237 pan troglod
Q95235 pongo pygma

34 235 12.5 390 1 DAF1_MOUSE
35 233.5 12.4 485 1 LEM2_BOVIN
36 227.5 12.1 331 1 FHR4_HUMAN
37 227 12.1 330 1 FHR3_HUMAN
38 222.5 11.9 330 1 FHR1_HUMAN
39 222.5 11.9 407 1 DAF2_MOUSE
40 222 11.8 646 1 LEM3_BOVIN
41 217 11.6 297 1 APOH_RAT
42 210 11.2 372 1 LEM1_MOUSE
43 203.5 10.8 372 1 LEM1_HUMAN
44 203.5 10.8 372 1 LEM1_PANTR
45 202.5 10.8 372 1 LEM1_PONPY


```
FT DISULFID 964 989 BY SIMILARITY.
FT DISULFID 994 1037 BY SIMILARITY.
FT DISULFID 1023 1048 BY SIMILARITY.
FT DISULFID 1053 1096 BY SIMILARITY.
FT DISULFID 1082 1107 BY SIMILARITY.
FT DISULFID 1114 1157 BY SIMILARITY.
FT DISULFID 1143 1168 BY SIMILARITY.
FT DISULFID 1172 1223 BY SIMILARITY.
FT DISULFID 1206 1233 BY SIMILARITY.
FT CARBOHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 773 773 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1225 1225 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1234 AA; 139082 MW; C5AC02F341B957F7 CRC64;

Query Match 67.2%; Score 1261; DB 1; Length 1234;
Best Local Similarity 65.7%; Pred. No. 8.3e-92;
Matches 216; Conservative 37; Mismatches 76; Indels 0; Gaps 0;

Qy 1 EDCNELPPRRNTEILTGSNDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLR 60
Db 19 EDCKGPPESEILGSWSQEQIYEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLR 78
Qy 61 KCKRPGCHGPGDTPGTFTLTGNNVFEYGVKAVYTCNMGYQLLGRINRECDTGTNDI 120
Db 79 ICRKPCGHGPGDTPGSRFVAVGSGFEGAKVYVTCDDGYQLLGRIDRECDGTNDI 138
Qy 121 PICEVVKCLPVTAPNGKIVSAMPEPDREYHFGQAVRVCNMGYKIEGDEEMHCSDDGFW 180
Db 139 PLCEVVKCLPVTLENGRIVSGAAETDQYRFGVVRFCNMGYKIEGKHEHCSGGLW 198
Qy 181 SKEPKCKVEISCKSPDVINGSPISOKIYKENERFQYKCNMGYSEBGDAVCTESGWRP 240
Db 199 SNEKPRCEILCTPRVNGDGINVKYKENERHYKCKHGYVPKRGDAVCTESGWS 258
Qy 241 LPSCEKSCNDNYPINGDYSPLRKHTGDEITYQCRNGFYPATRNTAKCTSTGTWIPAP 300
Db 259 QPCECEKRCSPYILNGIYVTHRIHRSDDIYRNCNMGYFVPTGTVSKCTPTGTWIPV 318
Qy 301 RCTLPCDYPDKHGLHNNRRYFPV 329
Db 319 RCTLPCDYPDKHGLHNNRRYFPV 347

RESULT 3
CPAH_BOVIN STANDARD; PRT; 685 AA.
ID Q28085;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Complement factor H (H factor 1) (Fragments).
GN HFI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 17-685 FROM N.A., AND SEQUENCE OF 1-16.
RC TISSUE=Liver;
RX MEDLINE=96202005; PubMed=8615824;
RA Soames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
RL in the interaction with complement component C3b.";
RL Biochem. J. 315:523-531(1996).
CC -!- FUNCTION: Factor H functions as a cofactor in the inactivation of
CC C3b by factor I and also increases the rate of dissociation of the
CC C3bb complex (C3 convertase) and the (C3b)NBB complex (C5
```

```
CC convertase) in the alternative complement pathway (By similarity).
CC -!- SIMILARITY: CONTAINS AT LEAST 13 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X98697; CAA67257.1; -.
CC HSP; P10998; 1VVD.
CC InterPro: IPR000436; Sushi_SCR_CCP.
CC Pfam: PF00084; sushi; 11.
CC SMART; SM00032; CCP; 11.
CC Complement alternate pathway; Plasma; Repeat; Sushi.
FT NON_TER 1 17
FT NON_CONS 16 17
FT DOMAIN <17 67 SUSHI 2.
FT DOMAIN 70 131 SUSHI 3.
FT DOMAIN 134 188 SUSHI 4.
FT DOMAIN 191 246 SUSHI 5.
FT DOMAIN 249 309 SUSHI 6.
FT DOMAIN 311 366 SUSHI 7.
FT DOMAIN 370 429 SUSHI 8.
FT DOMAIN 431 488 SUSHI 9.
FT DOMAIN 491 547 SUSHI 10.
FT DOMAIN 553 609 SUSHI 11.
FT DOMAIN 614 668 SUSHI 12.
FT DOMAIN 675 >685 SUSHI 13.
FT DISULFID 39 66 BY SIMILARITY.
FT DISULFID 71 117 BY SIMILARITY.
FT DISULFID 103 130 BY SIMILARITY.
FT DISULFID 135 176 BY SIMILARITY.
FT DISULFID 162 187 BY SIMILARITY.
FT DISULFID 192 234 BY SIMILARITY.
FT DISULFID 219 245 BY SIMILARITY.
FT DISULFID 250 297 BY SIMILARITY.
FT DISULFID 280 308 BY SIMILARITY.
FT DISULFID 312 354 BY SIMILARITY.
FT DISULFID 339 365 BY SIMILARITY.
FT DISULFID 371 417 BY SIMILARITY.
FT DISULFID 400 428 BY SIMILARITY.
FT DISULFID 432 476 BY SIMILARITY.
FT DISULFID 459 487 BY SIMILARITY.
FT DISULFID 492 534 BY SIMILARITY.
FT DISULFID 520 546 BY SIMILARITY.
FT DISULFID 554 597 BY SIMILARITY.
FT DISULFID 583 608 BY SIMILARITY.
FT DISULFID 615 656 BY SIMILARITY.
FT DISULFID 642 667 BY SIMILARITY.
FT NON_TER 685 685
SQ SEQUENCE 685 AA; 77536 MW; 69FC9DC8D530E872 CRC64;

Query Match 49.4%; Score 927.5; DB 1; Length 685;
Best Local Similarity 50.6%; Pred. No. 8.3e-66;
Matches 166; Conservative 35; Mismatches 68; Indels 59; Gaps 2;

Qy 1 EDCNELPPRRNTEILTGSNDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLR 60
Db 1 EDCKEPPPRKETEILSVS-----
Qy 61 KCKRPGCHGPGDTPGTFTLTGNNVFEYGVKAVYTCNMGYQLLGRINRECDTGTNDI 120
Db 19 -----GSPHLAEGNQFEGAKVYVTCDEGYQVGMENFRECDTGTNDI 63
Qy 121 PICEVVKCLPVTAPNGKIVSAMPEPDREYHFGQAVRVCNMGYKIEGDEEMHCSDDGFW 180
Db 64 PICEVVKCLPVTPENGKIFSDALEPDQYTYGVVQVFCNMGYKIEGKHEHCSAGGW 123
Qy 181 SKEPKCKVEISCKSPDVINGSPISOKIYKENERFQYKCNMGYSEBGDAVCTESGWRP 240
```

```

Db 124 SAETPKCVEIFKPPVILNGQAVLPKAYKQNERVOYRCAGFEQGRQDVTCTKSGWTP 183
QY 241 LPSCEKSCNDPIYPNGDYSPLRIKHRTGDEITYOCNRYFPATRGNTAKCTSTGWIPAP 300
Db 184 APTCEITCDPPRIPNGVYRPELSKYRGQDKITYECKGFFPEIRGTDATCTRDGWVPV 243
QY 301 RCTLKPDCYDPDIKHGGLYHENRRYFP 328
Db 244 RCWAKPCSYPIKHGRLYYS--YRGYFP 269

RESULT 4
C4BP_HUMAN STANDARD: PRT: 597 AA.
AC P04003;
DT 23-OCT-1986 (Rel. 02, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein alpha chain precursor (C4bp) (Proline-rich
DE protein) (PRP).
GN C4BPA OR C4BP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=90073699; PubMed=2590215;
RA Matsuguchi T., Okamura S., Aso T., Sata T., Niho Y.;
RT "Molecular cloning of the cDNA coding for proline-rich protein (PRP):
RT identity of PRP as C4b-binding protein.";
RL Biochem. Biophys. Res. Commun. 165:138-144(1989).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91113199; PubMed=1989602;
RA Aso T., Okamura S., Matsuguchi T., Sakamoto N., Sata T., Niho Y.;
RT "Genomic organization of the alpha chain of the human C4b-binding
RT protein gene.";
RL Biochem. Biophys. Res. Commun. 174:222-227(1991).
[3]
RP SEQUENCE OF 9-81 FROM N.A.
RX MEDLINE=88242821; PubMed=3378624;
RA Lintin S.J., Levin A.R., Reid K.B.M.;
RT "Derivation of the sequence of the signal peptide in human
RT C4b-binding protein and interspecies cross-hybridisation of the C4bp
RT cDNA sequence.";
RL FEBS Lett. 232:328-332(1988).
[4]
RP SEQUENCE OF 203-288 FROM N.A.
RX MEDLINE=86301119; PubMed=3017751;
RA Lintin S.J., Reid K.B.M.;
RT "Studies on the structure of the human C4b-binding protein gene.";
RL FEBS Lett. 204:77-81(1986).
[5]
RP SEQUENCE OF 80-597 FROM N.A.
RX MEDLINE=86025405; PubMed=3840370;
RA Chung L.P., Bentley D.R., Reid K.B.M.;
RT "Molecular cloning and characterization of the cDNA coding for C4b-
RT binding protein, a regulatory protein of the classical pathway of the
RT human complement system.";
RL Biochem. J. 230:133-141(1985).
[6]
RP SEQUENCE OF 49-88.
RX MEDLINE=85296001; PubMed=4033666;
RA Chung L.P., Gagnon J., Reid K.B.M.;
RT "Amino acid sequence studies of human C4b-binding protein: N-terminal
RT sequence analysis and alignment of the fragments produced by limited
RT proteolysis with chymotrypsin and the peptides produced by cyanogen
RT bromide treatment.";
RL Mol. Immunol. 22:427-435(1985).
[7]

```

```

RP ELECTRON MICROSCOPY, 3D-STRUCTURE, AND LIGAND BINDING.
RX MEDLINE=83221615; PubMed=6222381;
RA Dahlback B., Smith C.A., Mueller-Eberhard H.J.;
RT "Visualization of human C4b-binding protein and its complexes with
RT vitamin K-dependent protein S and complement protein C4b.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
CC AND WITH SERUM AMYLOID P COMPONENT.
CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC OF 3 POSSIBLE SORTS: A 570 kDa COMPLEX OF 7 ALPHA CHAINS AND 1
CC BETA CHAIN, A 530 kDa HOMOPOLYMER OF ALPHA CHAINS OR A 500 kDa
CC COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
CC THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
CC BINDING SITE FOR C4B AT THE END.
CC -1- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M31452; AAA36507.1; -
DR EMBL; M62486; AAA36506.1; -
DR EMBL; M62475; AAA36506.1; JOINED.
DR EMBL; M62476; AAA36506.1; JOINED.
DR EMBL; M62477; AAA36506.1; JOINED.
DR EMBL; M62478; AAA36506.1; JOINED.
DR EMBL; M62479; AAA36506.1; JOINED.
DR EMBL; M62480; AAA36506.1; JOINED.
DR EMBL; M62481; AAA36506.1; JOINED.
DR EMBL; M62482; AAA36506.1; JOINED.
DR EMBL; M62484; AAA36506.1; JOINED.
DR EMBL; M62485; AAA36506.1; JOINED.
DR EMBL; X07853; CAA30701.1; -
DR EMBL; X04284; CAA51244.1; -
DR EMBL; X04296; CAA27839.1; -
DR EMBL; X02865; CAA26617.1; -
DR PIR; A33588; NBHUC4.
DR HSSP; PI0998; LVVD.
DR MIM; I20830; -
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 8.
DR SMART; SM00032; CCP; 8.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
KW Polymorphism.
FT SIGNAL 1 48
FT CHAIN 49 597
FT DOMAIN 49 109
FT DOMAIN 112 171
FT DOMAIN 174 235
FT DOMAIN 238 295
FT DOMAIN 298 361
FT DOMAIN 364 423
FT DOMAIN 425 481
FT DOMAIN 483 539
FT DISULFID 50 96
FT DISULFID 81 108
FT DISULFID 113 154
FT DISULFID 140 170
FT DISULFID 175 217
FT DISULFID 203 234
FT DISULFID 239 281
C4B-BINDING PROTEIN ALPHA CHAIN.
SUSHI 1.
SUSHI 2.
SUSHI 3.
SUSHI 4.
SUSHI 5.
SUSHI 6.
SUSHI 7.
SUSHI 8.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.

```

```
FT DISULFID 267 294 BY SIMILARITY.
FT DISULFID 299 348 BY SIMILARITY.
FT DISULFID 332 360 BY SIMILARITY.
FT DISULFID 2364 2387 BY SIMILARITY.
FT DISULFID 2365 2409 BY SIMILARITY.
FT DISULFID 2399 422 BY SIMILARITY.
FT DISULFID 426 468 BY SIMILARITY.
FT DISULFID 454 480 BY SIMILARITY.
FT DISULFID 484 525 BY SIMILARITY.
FT DISULFID 511 538 BY SIMILARITY.
FT DISULFID 546 546 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 558 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .).
FT VARIANT 92 92 Q -> T.
FT VARIANT 357 357 /FTID-VAR_001977.
FT VARIANT 473 473 /FTID-VAR_001978.
FT VARIANT 473 473 W -> L (IN DBSNP:1801341).
FT SEQUENCE 597 AA: 67033 MW: 67E03F2EA8A516DD CRC64;

Query Match 18.8%; Score 353; DB 1; Length 597;
Best Local Similarity 29.3%; Pred. No. 1.3e-20;
Matches 100; Conservative 50; Mismatches 141; Indels 50; Gaps 21;

QY 20 SDQTYEGTQALYKCRPGV-RSLGNVIMVCRK-GEWVALNPLRKCKQKPCGHPGTPPGCT 77
DB 67 TETREKGTGLTKYCLPGVRSRSHSTQTLTNCSDGEWV-YNTFR-CIYKCRHPGE----- 118
QY 78 FTLTGNV-----FYGKAVVTCNEGYQLLGEINRECDTD-----GWTNDIPICEVVKC 128
DB 119 --LRNGQVEIKTDLFSGSGIERSCEGFLIGTTSR-CEVDRGVGWSHPLPQCEIVKC 175
QY 129 LPVTAPENKIVSSAMEPDREYHFGQAVRFVNSCYKIEGDEMCSDD----GFWSEK 184
DB 176 KPPDIRNGRSHGE-----ENFAYGFSVYSCDPFSLGLGHASISCTVENETIGVWRPSP 231
QY 185 PKCVIECKSPDVINGSPISQ-KIYKENERFQYKCNMGYEYSERGDV--C-TESGWRP 240
DB 232 PTCEKITCKPDSHGEMVSGFGPIYNYKDTIVFKCKQGFVL--RGSSVIHCDADSKNWP 289
QY 241 L-PSCEEKSCDN-PYIPNGDYS--PLRIK---HRTGDEITYCNRNGFYFAT-RGNTAKC- 291
DB 290 SPSPACEPNINLPDIPHASMETEYRPTKEDYVYVGVTLVRYCHPCYKPTTDEPTTVICQ 349
QY 292 TSTGWIAPRCTLKCDYDPIKHGGLYHENMRP-----YF 327
DB 350 KNLRTWYOGCEALCCPEPKLNGEITQHRKSRPANHCYVF 390
```

```
RESULT 5
CRL_HUMAN
ID CRL_HUMAN STANDARD; PRT; 2039 AA.
AC P17927;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement receptor type 1 precursor (C3b/C4b receptor) (CD35
DE antigen)
GN CRI OR C3BR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=69035992; PubMed=2972794;
RA Klickstein L.B., Bartow T.J., Miletic V., Rabson L.D., Smith J.A.,
FA Fearon D.T.;
RT *Identification of distinct C3b and C4b recognition sites in the
```

```
RT human C3b/C4b receptor (CRI, CD35) by deletion mutagenesis.*;
RL J. Exp. Med. 168:1699-1717(1988).
RN [2]
RP SEQUENCE OF 503-2039 FROM N.A.
RX MEDLINE=87168191; PubMed=2951479;
RA Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,
RA Fearon D.T.;
RT *Human C3b/C4b receptor (CRI). Demonstration of long homologous
RT repeating domains that are composed of the short consensus repeats
RT characteristics of C3/C4 binding proteins.*;
RL J. Exp. Med. 165:1095-1112(1987).
RN [3]
RP SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.
RX MEDLINE=86067975; PubMed=2933745;
RA Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;
RT *Identification of a partial cDNA clone for the human receptor for
RT complement fragments C3b/C4b.*;
RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).
CC FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUKOCYTES, GLOMERULAR
CC PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS, MEDIATES THE
CC BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE
CC ACTIVATED COMPLEMENT.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- POLYMORPHISM: CRI IS RESPONSIBLE FOR THE KNOX BLOOD GROUP SYSTEM.
CC -!- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A
CC LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SRCS OF LHR-A
CC CONTAINED A SITE DETERMINING C4 SPECIFICITY. AND THE N-TERMINAL
CC TWO SRCS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3
CC SPECIFICITY.
CC -!- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CRI.
CC -!- SIMILARITY: CONTAINS 30 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11569; AAA52297.1; -
DR EMBL; M11617; AAA52298.1; -
DR EMBL; M11618; AAA52299.1; -
DR EMBL; Y00816; CAA68755.1; -
DR EMBL; X05309; CAA28933.1; -
DR PIR; A28507; A28507.
DR PIR; A24748; A24748.
DR PIR; B24748; B24748.
DR PIR; C24748; C24748.
DR PIR; S03843; S03843.
DR HSP; P08603; 1HFI.
DR MIM; 120620; -
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 30.
DR SMART; SM00032; CCP; 30.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sushi; Blood group antigen.
FT SIGNAL 1 41
FT CHAIN 42 2039 COMPLEMENT RECEPTOR TYPE 1.
FT DOMAIN 42 1971 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1972 1996 POTENTIAL.
FT DOMAIN 1997 2039 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 42 42 PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
FT DOMAIN 42 100 SUSHI A1.
FT DOMAIN 103 162 SUSHI A2.
FT DOMAIN 165 233 SUSHI A3.
FT DOMAIN 237 294 SUSHI A4.
FT DOMAIN 296 354 SUSHI A5.
FT DOMAIN 357 417 SUSHI A6.
FT DOMAIN 420 488 SUSHI A7.
FT DOMAIN 492 550 SUSHI B1.
FT DOMAIN 553 612 SUSHI B2.
```

```
FT DOMAIN 615 683 SUSHI B3.
FT DOMAIN 687 744 SUSHI B4.
FT DOMAIN 746 804 SUSHI B5.
FT DOMAIN 807 867 SUSHI B6.
FT DOMAIN 870 938 SUSHI B7.
FT DOMAIN 942 1000 SUSHI C1.
FT DOMAIN 1003 1062 SUSHI C2.
FT DOMAIN 1065 1133 SUSHI C3.
FT DOMAIN 1137 1194 SUSHI C4.
FT DOMAIN 1196 1254 SUSHI C5.
FT DOMAIN 1257 1317 SUSHI C6.
FT DOMAIN 1320 1388 SUSHI C7.
FT DOMAIN 1395 1453 SUSHI D1.
FT DOMAIN 1456 1515 SUSHI D2.
FT DOMAIN 1518 1586 SUSHI D3.
FT DOMAIN 1590 1647 SUSHI D4.
FT DOMAIN 1649 1707 SUSHI D5.
FT DOMAIN 1710 1770 SUSHI D6.
FT DOMAIN 1773 1841 SUSHI D7.
FT DOMAIN 1847 1905 SUSHI E1.
FT DOMAIN 1908 1966 SUSHI E2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.
FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
FT DISULFID 1591 1633 BY SIMILARITY.
FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.

FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1739 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.
FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 959 959 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1028 1028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1310 1310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1481 1481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1540 1540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1605 1605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;

Query Match 18.7%; Score 350.5; DB 1; Length 2039;
Best Local Similarity 29.3%; Pred. No. 8.4e-20;
Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;

QY 3 CNE---LPPRRNTEILTGSWSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGWVALNPL 59
DB 43 CNAPELWLPARPNLT----DEFEFPGTYLNYECPGYSGRPFISICLNSVMTGARD- 97
QY 60 RKCQKRPCGHGDPFGTFTLTGGNVFEYGVKAVVYCNQGYQLLGEINYNREC---DTDG 115
DB 98 -RCRRKSCRNPDPVNGVHVYKNG--IQFGSQIKYSGCTKGYRLIGS--SSATCIISGDTVI 153
QY 116 WTNDIPICEVVKC-LPVTAPENGIYSSAMEPDREYHFGQAVRFVCNSG-----YKIEG 168
DB 154 WDNETPICDRIPCGLPPTI-TNGDFISTNRE---NEHYGSVVYTRCNPNGSGGRKRVELVG 209
QY 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDVINGSPIS-OKIYKENEREQYKCNMG 222
DB 210 EPSIYCTSDNDQVGIIWSGPAQCIIIPNCTPPNVENGILVSDNRSLFSLNEVVEFRCPQG 269
QY 223 YEVSERGDVCTE-SGWRP-LPSCCEKSCDNPVYPN-----GDYSPLRKHRTG 269
DB 270 FVMKGPVRVKCALNKNWEPFLSC-SRVCQPP--PDVLAERTQDKDNFSP-----G 319
QY 270 DEITYQCRNGFYPATRGNTAKTSTG-WIP-APRCTLKPCD 308
DB 320 QEVFYSCEPG-YDLRGAASMRCTPQGDWSPAAPTCEVKSCD 359

RESULT 6
C4BP-BOVIN
ID C4BP-BOVIN STANDARD; PRT; 610 AA.
AC Q28065;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein alpha chain precursor (C4bp).
GN C4BPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
```



```
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95015909; PubMed=7930621;
RA Hillarp A., Thern A., Dahlbaeck B.;
RT "Bovine C4b binding protein. Molecular cloning of the alpha- and
beta-chains provides structural background for lack of complex
formation with protein S.";
RL J. Immunol. 153:4190-4199(1994).
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC (BY SIMILARITY).
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z31693; CAA83498.1; -
CC HSSP; P10998; 1VVD.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00084; sushi; 8.
CC SMART; SM00032; CCP; 8.
CC Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT SIGNAL 1 48
FT CHAIN 49 610
FT DOMAIN 49 108
FT DOMAIN 111 170
FT DOMAIN 173 235
FT DOMAIN 238 295
FT DOMAIN 298 363
FT DOMAIN 366 426
FT DOMAIN 428 484
FT DOMAIN 486 542
FT DISULFID 50 95
FT DISULFID 80 107
FT DISULFID 112 153
FT DISULFID 139 169
FT DISULFID 174 217
FT DISULFID 203 234
FT DISULFID 239 281
FT DISULFID 267 294
FT DISULFID 299 350
FT DISULFID 334 362
FT DISULFID 366 7390
FT DISULFID 7367 7412
FT DISULFID 7402 425
FT DISULFID 429 471
FT DISULFID 457 483
FT DISULFID 487 528
FT DISULFID 514 541
FT DISULFID 549 549
FT DISULFID 561
FT CARBOHYD 66 66
FT CARBOHYD 221 221
FT CARBOHYD 525 525
FT CARBOHYD 602 602
FT SEQUENCE 610 AA; 58886 MW; D806B270E8A0558 CRC64;
Query Match 17.1%; Score 320.5; DB 1; Length 610;
Best Local Similarity 27.7%; Pred. No. 5e-18;
```

```
Matches 99; Conservative 54; Mismatches 142; Indels 63; Gaps 22;
QY 6 LPPRRNTEILTGSWSQTYPEGTQAIYKCRPGYR--SLGNVIMVCRKGEWALNPLRKQ 63
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
52 IPPYLDFAFFINELNETRFETGTLRTCRPGYRISRRKNFLICDGTDMW---KYKEFCV 108
QY 64 KRPGCHPGDTPFGFTLTGGNV-----FEYGVKAVYTCNEGYQLLGEIN-YRECDT 115
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
109 KKRCENPGE-----LLNGQVIVKTDYSFGSETEFSCSEGYVLIGSANSYQLQDKGV 161
QY 116 WTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRYHFQGVQAVFVNCNYSKTEGDEE 175
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
162 WSDPLPQCIITAKCEPPPTISNGR--HNGGDED-EYTYGSSVYSCDRDFSLGKASISCR 218
QY 176 DD-----GFWSKEKPKVEISCKSPDIVNGSPISOKI-----IYKENERFOYK 225
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
219 VENKTIGWSPSPSCKKVICQPVVDG-----KITSFGPGPIYTYQOSIYVACNKG 273
QY 226 SERGDV--C-TESGWR-PLPSCEEKSC-DNPYIPNGDISPLRIKHKRTGDE----- 271
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
274 --EGDSLIIHCEADNSNMPPPTCELNGCLGLPHIPHALWE--RYDHOTQTTEQVYD 329
QY 272 ITVQCRNGFVPATRG-NTAKCTST-GWIPAPRCTLKPCDYPDIKHGG--LYHENMR 325
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
330 LSYKCHFGYKPETDGPTTVCQSNLEWSPYIECKVEYCCPEPNNYGSITLH---RRP 384
RESULT 7
C4BP_RAT STANDARD; PRT; 558 AA.
ID C4BP_RAT
AC Q63514;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein alpha chain precursor (C4bp).
GN C4BPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=97166082; PubMed=9013975;
RA Hillarp A., Wiklund H., Thern A., Dahlback B.;
RT "Molecular cloning of rat C4b binding protein alpha- and beta-chains:
structural and functional relationships among human, bovine, rabbit,
mouse, and rat proteins.";
RL J. Immunol. 158:1315-1323(1997).
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
CC AND WITH SERUM AMYLOID P COMPONENT.
CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z50051; CAA90391.1; -
CC HSSP; P10998; 1VVD.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00084; sushi; 8.
CC SMART; SM00032; CCP; 8.
```

KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT SIGNAL 1 13 BY SIMILARITY.
FT CHAIN 14 558 C4B-BINDING PROTEIN ALPHA CHAIN.
FT DOMAIN 14 73 SUSHI 1.
FT DOMAIN 14 73 SUSHI 2.
FT DOMAIN 138 200 SUSHI 3.
FT DOMAIN 203 259 SUSHI 4.
FT DOMAIN 262 325 SUSHI 5.
FT DOMAIN 328 387 SUSHI 6.
FT DOMAIN 389 444 SUSHI 7.
FT DOMAIN 446 502 SUSHI 8.
FT DISULFID 15 60 BY SIMILARITY.
FT DISULFID 45 72 BY SIMILARITY.
FT DISULFID 77 118 BY SIMILARITY.
FT DISULFID 104 134 BY SIMILARITY.
FT DISULFID 139 182 BY SIMILARITY.
FT DISULFID 168 199 BY SIMILARITY.
FT DISULFID 204 246 BY SIMILARITY.
FT DISULFID 232 258 BY SIMILARITY.
FT DISULFID 263 312 BY SIMILARITY.
FT DISULFID 296 324 BY SIMILARITY.
FT DISULFID 7328 7351 BY SIMILARITY.
FT DISULFID 7329 7373 BY SIMILARITY.
FT DISULFID 7363 386 BY SIMILARITY.
FT DISULFID 390 431 BY SIMILARITY.
FT DISULFID 417 443 BY SIMILARITY.
FT DISULFID 447 488 BY SIMILARITY.
FT DISULFID 474 501 BY SIMILARITY.
FT DISULFID 509 509 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 521 521 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 31 31 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 558 AA; 62266 MW; 592F0C667EDIE5FF CRC64;

Query Match 16.9%; Score 317; DB 1; Length 558;
Best Local Similarity 27.0%; Pred. No. 8.6e-18;
Matches 110; Conservative 46; Mismatches 128; Indels 124; Gaps 24;
Qy 7 PPRNTEILTGSWSQDT-YPEQTOAIYKCRPGY-RSLGNVIMVCRK-GEW---VALNPLR 60
Db 17 PPPDLPALPASEMKNQDTFESHHTLRYNCRPGYSRASSQSLYKPLGKQWQINIA----- 71
Qy 61 KCKRPGCHGDPPTGTTLTGGNV-----FEYGVKAVYTCNEGQQLLG-EINYRECDTD 114
Db 72 -CVKRCRNPGD-----LQNGKVEVKTDFLFGSQIEFSCSEGVILIGSSTSYCEIOGK 123
Qy 115 G-WTNDIPICEVVKCLPVTAPENCK-----IVSSAMEPD----- 147
Db 124 GVSWSDPLECVIAKCGMPDLSNCKHNGREEEFTYRSVYTKCDPDTLLGNASITCT 183
Qy 148 -----REYHFGQAVRFVNCYKIEGDDE 171
Db 184 VVNTGVWSPSPPCERIIICPWPVKVLHGTINSFKHYKYKDSVRFQCGFVLURSGV 243
Qy 172 MHCSDGFWSEKPKVEISCKS-PDVINGSPIS-----QKIYKENERFOYKCMNGY 225
Db 244 IHCADGWS-PVPVCELNSCTDIPDIPNAALITSPRKRKEDVYPVGVTLVYICRPGYEP 302
Qy 226 SERG-DAVCTES-GWRPLPSCSEKSCNDVPINGDYSPLR-IKHRT-----GD 270
Db 303 ATROPMTVICQKDLWSMLRCKEICCPVP-----DPKSVRVIQHEKAHPNDNDCTYFEGD 357
Qy 271 EITYQCRNGFYPATRGNTAKTSTG-WIP-APRCTLKPDPYD-IKHG 315
Db 358 EVSYTCQNDIN-----LTATCKSDGTHWPTPTPSCH-QSCDFPPAIAGH 399

RESULT 8
C4BP_MOUSE

C4BP_MOUSE STANDARD; PRT; 469 AA.
P08607;
01-AUG-1988 (Rel. 08, Created)
01-FEB-1994 (Rel. 28, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
C4B-binding protein precursor (C4bp).
C4BPA OR C4BP.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
MEDLINE=88024997; PubMed=3663616;
Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;
"cDNA structure of murine C4b-binding protein, a regulatory component
of the serum complement system";
Biochemistry 26:4668-4674(1987).
-|- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
(C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
ALSO ACCELERATES THE DEGRADATION OF THE C4B/C2A COMPLEX (C3
CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
-|- SUBUNIT: HOMOHEPTAMER; NOT COVALENTLY LINKED. MOUSE LACKS THE
BETA CHAIN OF C4BP.
-|- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
-|- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS
-|- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M17122; AAA37312.1; ALT_INIT.
PIR; A27117; NBMSC4.
HSSP; P10998; 1VVD.
MGD; MGI:88229; C4bp.
InterPro; IPR000436; Sushi_SCR_CCP.
Pfam; PF00084; sushi; 6.
SMART; SM00032; CCP; 6.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT SIGNAL 1 56
FT CHAIN 57 469 C4B-BINDING PROTEIN.
FT DOMAIN 57 116 SUSHI 1.
FT DOMAIN 119 177 SUSHI 2.
FT DOMAIN 180 241 SUSHI 3.
FT DOMAIN 244 300 SUSHI 4.
FT DOMAIN 302 356 SUSHI 5.
FT DOMAIN 358 414 SUSHI 6.
FT DISULFID 58 103 BY SIMILARITY.
FT DISULFID 88 115 BY SIMILARITY.
FT DISULFID 120 160 BY SIMILARITY.
FT DISULFID 146 176 BY SIMILARITY.
FT DISULFID 181 223 BY SIMILARITY.
FT DISULFID 209 240 BY SIMILARITY.
FT DISULFID 245 287 BY SIMILARITY.
FT DISULFID 273 299 BY SIMILARITY.
FT DISULFID 303 343 BY SIMILARITY.
FT DISULFID 329 355 BY SIMILARITY.
FT DISULFID 359 400 BY SIMILARITY.
FT DISULFID 386 413 BY SIMILARITY.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 428 428 N-LINKED (GLCNAC. . .) (POTENTIAL).

RA Grundmann U., Nerlich C., Rein T., Zettlmeissl G.;
 FT "Complete cDNA sequence encoding the B subunit of human factor XIII.";
 RL Nucleic Acids Res. 18:2817-2817(1990).
 RN [5]
 RP VARIANT PHE-450.
 RX MEDLINE-93313189; PubMed-8324218;
 RA Hashiguchi T., Saito M., Morishita E., Matsuda T., Ichinose A.;
 RT "Two genetic defects in a patient with complete deficiency of the b-
 subunit for coagulation factor XIII.";
 RL Blood 82:145-150(1993).
 CC -!- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,
 BUT IS THOUGHT TO STABILIZE THE A SUBUNITS AND REGULATE THE RATE
 OF TRANSLUTAMINASE FORMATION BY THROMBIN.
 CC -!- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.
 CC -!- DISEASE: A DEFICIENCY IN FAL3 CAN RESULT IN A LIFE-LONG BLEEDING
 TENDENCY, DEFECTIVE WOUND HEALING, AND HABITUAL ABORTION.
 CC -!- SIMILARITY: CONTAINS 10 SUSHI (SCR) DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M64554; AAA51821.1; ALT_SEQ.
 DR EMBL; M14057; AAA88042.1; -
 DR EMBL; X51823; CAA36123.1; -
 DR PIR; A23830; A23830.
 DR PIR; A36397; A36397.
 DR PIR; S09980; S09980.
 DR HSP; P08603; 1HFI.
 DR MIM; 134580; -
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00084; sushi; 8.
 DR SMART; SM00032; CCP; 8.
 KW Transferase; Plasma; Blood coagulation; Repeat; Glycoprotein; Signal;
 KW Sushi; Disease mutation.
 FT SIGNAL 1 20
 FT CHAIN 21 661 COAGULATION FACTOR XIII B CHAIN.
 FT DOMAIN 24 88
 FT DOMAIN 90 147
 FT DOMAIN 152 209
 FT DOMAIN 212 268
 FT DOMAIN 273 328
 FT DOMAIN 335 390
 FT DOMAIN 395 451
 FT DOMAIN 453 516
 FT DOMAIN 523 579
 FT DOMAIN 581 647
 FT DISULFID 25 76 BY SIMILARITY.
 FT DISULFID 59 87 BY SIMILARITY.
 FT DISULFID 91 135 BY SIMILARITY.
 FT DISULFID 118 146 BY SIMILARITY.
 FT DISULFID 153 197 BY SIMILARITY.
 FT DISULFID 180 208 BY SIMILARITY.
 FT DISULFID 213 255 BY SIMILARITY.
 FT DISULFID 241 267 BY SIMILARITY.
 FT DISULFID 274 316 BY SIMILARITY.
 FT DISULFID 302 327 BY SIMILARITY.
 FT DISULFID 336 378 BY SIMILARITY.
 FT DISULFID 364 389 BY SIMILARITY.
 FT DISULFID 396 439 BY SIMILARITY.
 FT DISULFID 425 450 BY SIMILARITY.
 FT DISULFID 454 505 BY SIMILARITY.
 FT DISULFID 486 515 BY SIMILARITY.
 FT DISULFID 524 567 BY SIMILARITY.
 FT DISULFID 553 578 BY SIMILARITY.
 FT DISULFID 582 636 BY SIMILARITY.
 FT DISULFID 616 646 BY SIMILARITY.
 FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 545 545 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT 545

FT SITE 617 619 CELL ATTACHMENT SITE.
 FT VARIANT 450 450 C -> F (IN F13B DEFICIENCY).
 FT SEQUENCE 661 AA: 75491 MW: 57A2FB46560857F2 CRG64;
 SQ
 Query Match 16.2%; Score 303; DB 1; Length 661;
 Best Local Similarity 26.4%; Pred. No. 1.3e-16;
 Matches 93; Conservative 49; Mismatches 136; Indels 74; Gaps 20;
 QY 13 EILTGSWS--DQTYPEGTQAIYKRGYSLGNVIMVCRKGWV-----ALNPLRKQC 63
 DB 157 ELYNGNYSYTTQTFKVKQVOCATGYTAGG-----KTEVECLTYGWSLTP--KCT 209
 QY 64 KRPGC-----HPGDTPEFTTTLTGGNVFEYGVKAVYTCNEGYQLLGEINRECDT 113
 DB 210 KLKCSLSRLIENGYFHPVK-----QTVEGDVVQFFCHENYVLSGS-DLIQCVN 257
 QY 114 DQWTDNDIPICE--VVKCLPVTAPENCKIYSSAMEPDREYHFGQAVFVCSNGYKIGDEE 171
 DB 258 FGMYPESPVCERRRRCPPPLPINSKIQTHS-----TTRHGEIVHIEGELNFEIHGSAE 313
 QY 172 MHCSDDDGFWSEKPKCQE----ISCKSPDVI--NGSPISQKIYKENERFYKCNMGVEYS 226
 DB 314 IRC-EDGKWT-EPPKCIQEQKVAEEPPPIENGANLHSHKIYYNGDKVYACKSGILLH 371
 QY 227 ERGDAVCTESGWRPLPSCDE--KSCDN-PYIPNGDYSP--LRKHRTGDEITTYOCNRGFP 282
 DB 372 GSNEITCNRGKWTLPPECVENNENCKHPVVMGAVDGLASYATGSSVEYRC-NEYVL 430
 QY 283 ATRGNVAKTSTGTWIPAPRCTLKPC-----DYPD-----IKHGGL 317
 DB 431 LRGSKISRCEQGWSSPPVC--LEPCTVNVDMNRMNIEMKWKYEGKVLHGDL 481
 RESULT 12
 VCP_VACCV
 ID VCP_VACCV STANDARD; PRT; 263 AA.
 AC P10998;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Complement control protein precursor (VCP) (Secretory protein 35)
 DE (Protein C3) (28 kDa protein).
 GN C3L.
 OS Vaccinia virus (strain WR), and
 OS Vaccinia virus (strain Copenhagen).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=10254, 10249;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.
 RC STRAIN-WR;
 RX MEDLINE=88318974; PubMed=3412473;
 RA Kotwal G.J., Moss B.;
 RT "Vaccinia virus encodes a secretory polypeptide structurally related
 to complement control proteins.";
 RL Nature 335:176-178(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WR;
 RX MEDLINE=89073756; PubMed=2849238;
 RA Kotwal G.J., Moss B.;
 RT "Analysis of a large cluster of nonessential genes deleted from a
 vaccinia virus terminal transposition mutant.";
 RL Virology 167:524-537(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COPENHAGEN;
 RX MEDLINE=91021027; PubMed=2219722;
 RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
 RA Paoletti E.;
 RT "The complete DNA sequence of vaccinia virus.";

[illegible]

```

FT DOMAIN 453 516 SUSHI 8.
FT DOMAIN 523 579 SUSHI 9.
FT DOMAIN 581 647 SUSHI 10.
FT DISULFID 25 76 BY SIMILARITY.
FT DISULFID 59 87 BY SIMILARITY.
FT DISULFID 91 135 BY SIMILARITY.
FT DISULFID 118 146 BY SIMILARITY.
FT DISULFID 153 197 BY SIMILARITY.
FT DISULFID 180 208 BY SIMILARITY.
FT DISULFID 213 255 BY SIMILARITY.
FT DISULFID 241 267 BY SIMILARITY.
FT DISULFID 274 316 BY SIMILARITY.
FT DISULFID 302 327 BY SIMILARITY.
FT DISULFID 336 378 BY SIMILARITY.
FT DISULFID 364 389 BY SIMILARITY.
FT DISULFID 396 439 BY SIMILARITY.
FT DISULFID 425 450 BY SIMILARITY.
FT DISULFID 454 505 BY SIMILARITY.
FT DISULFID 486 515 BY SIMILARITY.
FT DISULFID 524 567 BY SIMILARITY.
FT DISULFID 553 578 BY SIMILARITY.
FT DISULFID 582 636 BY SIMILARITY.
FT DISULFID 616 646 BY SIMILARITY.
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 545 545 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 668 AA; 76078 MW; 80BC9E00A9E53FA6 CRC64;

Query Match 15.78; Score 294; DB 1; Length 668;
Best Local Similarity 28.18; Pred. No. 6.8e-16;
Matches 93; Conservative 41; Mismatches 141; Indels 56; Gaps 21;

QY 13 ELLTGSWS--DQYPEGTAIYKCRPGY-----RSLGNVIVCRKGEWALNPLRKCOKR 65
Db 157 ELEHGYNTORTFKVDIVATCTAGYTTTGKOTGEA--EQQANGW-SLTP--QCNKL 211
QY 66 PCG-----HPGDFPFGTFTLTGNVFEYGVKAVYTCNEGYOLLGEINRYRCDTDG 115
Db 212 MCSSLRLIENGYFHPVK-----QTYEGDLVQFFCHENYLSGS-DLIQCYNFG 259
QY 116 WTNDIPICE--VWKLPLVTAPENKIVSSAMEP-DREYHFGQAVRFVNSGYKIEGDREM 172
Db 260 WYPSPICEGRNRNCPPPVPLNSKI-----QPHSTTYRHGERVHIECELNFVIQGSSEL 314
QY 173 HCSDDGFWSEKPKCQE---ISCKS-PDVINGSPISQIIYKENERFYKCNMGYEYSE 227
Db 315 LC-ENGKW-TFPPKIEKEKVAQPPSPVNGVAPHSEIYSGDKVYTCRGGGYSLRG 372
QY 228 RGDVAVCTESGWRPLPSCBE--KSC-DNPIYIPNG-DYSLPLRIKHRTGDEITYQCRNGFYPA 283
Db 373 SSTITCNRGRWTLPEECVENIENCKPPDPIANGVVDGLLASYTTGSSVEYRC-NEYVLL 431
QY 284 TRGNVAKTSTGWPAPRCTLKPCDYPDIKH 314
Db 432 KGSETSRCEQAGWSSPPVC-LEPCTI-DVDH 460

RESULT 14
LEM3_HUMAN
ID LEM3_HUMAN STANDARD; PRT; 830 AA.
AC P16109;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE P-selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM)
DE (CD62P) (leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN SELP OR GMRP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE-89168432; PubMed-2466574;
RA Johnston G.I., Cook R.G., McEver R.P.;
RT "Cloning of GMP-140, a granule membrane protein of platelets and
endothelium: sequence similarity to proteins involved in cell
adhesion and inflammation.";
RL Cell 56:1033-1044(1989).
RN [2]
RP PALMITOYLATION.
RX MEDLINE-93266599; PubMed-7684381;
RA Fujimoto T., Stroud E., Whalley R.E., Prescott S.M., Muszbek L.,
Laposata M., McEver R.P.;
RT "P-selectin is acylated with palmitic acid and stearic acid at
cysteine 766 through a thioester linkage.";
RL J. Biol. Chem. 268:11394-11400(1993).
RN [3]
RP STRUCTURE BY NMR OF 160-199.
RX MEDLINE-97057176; PubMed-8901515;
RA Freedman S.J., Sanford D.G., Bachovchin W.W., Furie B.C., Baleja J.D.,
Furie B.;
RT "Structure and function of the epidermal growth factor domain of P-
selectin.";
RL Biochemistry 35:13733-13744(1996).
RN [4]
RP 3D-STRUCTURE MODELING OF 42-161.
RX MEDLINE-94093388; PubMed-7505680;
RA Bajorath J., Stenkamp R., Aruffo A.;
RT "Knowledge-based model building of proteins: concepts and examples.";
RL Protein Sci. 2:1798-1810(1993).
RN [5]
RP VARIANTS ASN-331; ASP-603; VAL-640 AND PRO-756.
RX MEDLINE-98334547; PubMed-9668170;
RA Heurmann S.M., Ricard S., Nicaud V., Mallet C., Evans A.,
Ruidavets J.B., Arviller D., Luc G., Cambien F.;
RT "The P-selectin gene is highly polymorphic: reduced frequency of the
Pro715 allele carriers in patients with myocardial infarction.";
RL Hum. Mol. Genet. 7:1277-1284(1998).
CC -1- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO
THE CELL SURFACE.
CC -1- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 9 SUSHI (SCR) DOMAINS.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD62P entry;
WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd62p.htm".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M60234; AAA35910.1; JOINED.
CC EMBL; M60217; AAA35910.1; JOINED.
CC EMBL; M60218; AAA35910.1; JOINED.
CC EMBL; M60219; AAA35910.1; JOINED.
CC EMBL; M60222; AAA35910.1; JOINED.
CC EMBL; M60223; AAA35910.1; JOINED.
CC EMBL; M60224; AAA35910.1; JOINED.
CC EMBL; M60225; AAA35910.1; JOINED.
CC EMBL; M60226; AAA35910.1; JOINED.
CC EMBL; M60227; AAA35910.1; JOINED.
CC EMBL; M60228; AAA35910.1; JOINED.
CC EMBL; M60229; AAA35910.1; JOINED.
CC EMBL; M60231; AAA35910.1; JOINED.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:11:38 ; Search time 67.22 Seconds
(without alignments)

Title: US-09-316-163-11
 Perfect score: 1876
 Sequence: 1 EDCNELPPRRVTEILTGSWS.....PDIKHGGLYHENMRPPYFPV 329

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database :
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeo.*
```

SUMMARY

Result No.	Query %			Length	DB	ID	Description
	Score	Match					
1	1876	100.0	1172	4	Q9NU87	Q9NU87	homo sapien
2	1366	72.8	1236	11	Q91YB6	Q91YB6	rattus norv
3	903	48.1	669	6	Q28085	Q28085	bos taurus
4	486	25.9	1053	13	Q91Z75	Q91Z75	paralabrax
5	377	20.1	522	6	Q28769	Q28769	papio cynoc
6	356.5	19.0	3567	11	Q9ES77	Q9ES77	mus musculus
7	353	18.8	679	11	Q99254	Q99254	mus musculus
8	351	18.7	1911	6	Q29528	Q29528	papio hamad
9	350.5	18.7	559	4	Q9UQV2	Q9UQV2	homo sapien
10	350.5	18.7	2039	4	Q16745	Q16745	homo sapien
11	350.5	18.7	2489	4	Q16744	Q16744	homo sapien
12	349.5	18.6	645	12	Q9WRU2	Q9WRU2	macaca mula
13	348	18.5	2014	6	Q29530	Q29530	pan troglod
14	342.5	18.3	661	6	Q29531	Q29531	pan troglod
15	339	18.1	522	6	Q28797	Q28797	pan troglod
16	337.5	18.0	560	5	Q22328	Q22328	caenorhabdi

ALIGNMENTS

```

RESULT      1
QIDN087    PRELIMINARY;          PRT: 1172 AA.
AC Q9NU87;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2000 (TrEMBLrel. 19, Last annotation update)
DE DJI77P10.1.1 (H FACTOR 1 (COMPLEMENT) ISOFORM 1).
.N HFI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL049744; CAB70597.1; -.
DR HSSP; P08603; IHFH.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 19.
DR SMART; SM00032; CCP; 19.
SQ SEQUENCE 1172 AA; 132087 MW; 8FB954C4B4FA54 CRC64;

Query Match           100.0%; Score 1876; DB 4; Length 1172;
Best Local Similarity 100.0%; Pred. No. 1.4e-165;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDNCNELPPRRNTILGNSDQTYPGTQAIIKCRPGYSLSGNVIMVCRKGHWALNP 60
   |||
Db 19 EDNCNELPPRRNTILGNSDQTYPGTQAIIKCRPGYSLSGNVIMVCRKGHWALNP 78
   |||

QY 61 KCKRRCGHPGDTPFGFTLTGGNVFEYGKAVYTTCNEG YQLLGELINRECDTGWTNDI 120
   |||
Db 79 KCKRRCGHPGDTPFGFTLTGGNVFEYGKAVYTTCNEG YQLLGELINRECDTGWTNDI 138
   |||

QY 121 PTCEVVKLPTVPENGGKIIVSSAMEPDREYHFQGAVRFYCNSGYIKEGDEEMHCSDDGFW 180
   |||

```

```
Db 139 PICEVVKLPVTAPENGKIVSSAMEPDREYHFGQAVFVNCNSGYKIEGDEMHCSDDGF 198
QY 181 SKEKPKCIVEISCKSPDVINGSPISQKIYKENERFOYKCNMGYSEYERGDVAVCTESGWRP 240
Db 199 SKEKPKCIVEISCKSPDVINGSPISQKIYKENERFOYKCNMGYSEYERGDVAVCTESGWRP 258
QY 241 LPSCEKSCDNPYPNGDYSPLRIKHRTGDEITYQCRNGFYPATRGNTAKCTSTGWIAPAP 300
Db 259 LPSCEKSCDNPYPNGDYSPLRIKHRTGDEITYQCRNGFYPATRGNTAKCTSTGWIAPAP 318
QY 301 RCTLKPCDYPDKHGLYHNNRRPFPV 329
Db 319 RCTLKPCDYPDKHGLYHNNRRPFPV 347

RESULT 2
QY1YB6
ID QY1YB6 PRELIMINARY: PRT: 1236 AA.
AC QY1YB6;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE COMPLEMENT INHIBITORY FACTOR H.
GN FH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;
RA Demberg T., Goetze O., Schlaf G.;
RT "Rat complement factor H: molecular cloning, sequencing and expression
in tissues and isolated cells.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ320522; CAC67513.1; -.
SQ SEQUENCE 1236 AA; 140343 MW; 1AC89FFA28232BEF CRC64;

Query Match 72.8%; Score 1366; DB 11; Length 1236;
Best Local Similarity 71.1%; Pred. No. 3.8e-118;
Matches 234; Conservative 31; Mismatches 64; Indels 0; Gaps 0;

QY 1 EDCNELPPRRNTIELTGSWSDQTYPEGTQAIYKRCGYSLGNVIMVCRKGWVALNPLR 60
Db 19 EDCKGGPPRENSILLSGNSEQLYSEGQATYKRCGYRTLTGTVKVKNGEWVPSNPSR 78
QY 61 KCOKRPGHGDTPFGTFTLTGNNVFEYGVKAVYTCNEGYYQLLGEINRYRECDTGTNDI 120
Db 79 ICKKRPGHGDTPFGSFLAVGSEFEFGAKVYVTCDEGYQLLGEIDYRECDADGWTNDI 138
QY 121 PICEVVKLPVTAPENGKIVSSAMEPDREYHFGQAVFVNCNSGYKIEGDEMHCSDDGF 180
Db 139 PICEVVKLPVTELENGRIVSGAAEPDQYIFGVGVFECNSGFKIEGOKEMHCSENGLW 198
QY 181 SKEKPKCIVEISCKSPDVINGSPISQKIYKENERFOYKCNMGYSEYERGDVAVCTESGWRP 240
Db 199 SNEKPOCIVEISCLPPRVENGDIYLPVYKENERFOYKCKQGFVYKERGDVAVCTGSGWRP 258
QY 241 LPSCEKSCDNPYPNGDYSPLRIKHRTGDEITYQCRNGFYPATRGNTAKCTSTGWIAPAP 300
Db 259 QPSCEEMCLTYPINGDIYTPRIKHRIIDDEIRYCKNGLYPATRSPVSKCTITGWIAPAP 318
QY 301 RCTLKPCDYPDKHGLYHNNRRPFPV 329
Db 319 RCTLKPCDYPDKHGLYHNNRRPFPV 347

RESULT 3
QY2805
ID QY2805 PRELIMINARY: PRT: 669 AA.
AC QY2805;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
```

```
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=96202005; PubMed=8615824;
RA Soames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
in the interaction with complement component C3b.";
RL Biochem. J. 315:523-531(1996).
DR EMBL; X98697; CAA67257.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 11.
DR SMART; SM00032; CCP; 11.
FT NON_TER 1
FT NON_TER 669
SQ SEQUENCE 669 AA; 75683 MW; D0D9DB30EE747AC2 CRC64;

Query Match 48.1%; Score 903; DB 6; Length 669;
Best Local Similarity 60.9%; Pred. No. 1.9e-75;
Matches 154; Conservative 33; Mismatches 64; Indels 2; Gaps 1;

QY 76 GTFTLTGNNVFEYGVKAVYTCNEGYYQLLGEINRYRECDTGTNDIPICEVVKLPVTAPE 135
Db 3 GSPHLAEGNQFEYGAQVYVTCDEGYQWGMNFRECDTGTNDIPICEVVKLPVTEPE 62
QY 136 NKKIVSSAMEPDREYHFGQAVFVNCNSGYKIEGDEMHCSDDGFVSKERPKCIVEISCKSP 195
Db 63 NGKIFSDALEPDQYTYGVVQVFEVFEVFEVFEVFEVFEVFEVFEVFEVFEVFEVFEVFEV 122
QY 196 DVINGSPISQKIYKENERFOYKCNMGYSEYERGDVAVCTESGWRPLPSCPEKSCDNPYP 255
Db 123 VILNGQAVLPKATYKQNERVQVRCAGFEYQGRGDTVCTKSGWTAPTCLIEITCDPPRIP 182
QY 256 NGDYSPLRIKHRTGDEITYQCRNGFYPATRGNTAKCTSTGTWIPAPRCTTLKPCDYPDIKHG 315
Db 183 NGVYRPELSKYRQDKITVECKGFFPEIRGTATCTRDGQVVPVPCAMKPCSPYVIKHG 242
QY 316 GLYHNNRRPFPV 328
Db 243 RLYYS--YRGYFP 253

RESULT 4
QY1275
ID QY1275 PRELIMINARY: PRT: 1053 AA.
AC QY1275;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE COMPLEMENT REGULATORY PLASMA PROTEIN.
OS Parabrax nebulifer (barred sand bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Percoidae;
OC Serranidae; Paralabrax.
OX NCBI_TaxID=30873;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=94318039; PubMed=8042982;
RA Dahmen A., Kaidon T., Zipfel P.F., Gigli I.;
RT "Cloning and characterization of a cDNA representing a putative
complement-regulatory plasma protein from barred sand bass (Parabrax
nebulifer).";
```

```
RL Biochem. J. 301:391-397(1994).
DR EMBL; L21703; AAA92556.1; -.
DR HSSP; P08603; LHFH.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 16.
DR SMART; SM00032; CCP; 16.
SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D5D3 CRC64;

Query Match      25.9%; Score 486; DB 13; Length 1053;
Best Local Similarity 29.7%; Pred. No. 1.8e-36;
Matches 114; Conservative 48; Mismatches 134; Indels 88; Gaps 13;

QY 21 DQTPPEGTQAIYKCRGYSRLGNVIMVCRKGEWVALNPLRKKCRKCGHGGDTPFGTFL 80
DB 47 EASYPGGVRVCGNVGYS--GFFKLVCVEGKWETRG--AKQPRSCGHGPDQAQADFHL 102
QY 81 TGNVFEYGVKAVYTCNEGQYLLGEINRECDTGTNDIPICEVVKCLPVTAPENGKIV 140
DB 103 AEGNDFVFGSKVYTCQKGYQWYSRINRYRCVAEGHDGVVPCESQOCPLIHVDNNOVVI 162
QY 141 SSAMEPDREYHFGQAVFVNCNSYKI--EGDEEMHCSDDGFWSEKPKVCVEISCKSPDVIN 199
DB 163 GG----PEEATFCNVVRFSCKSRSEILDGSPELYCDERGDWSGVPVKRAITCAIPIEN 218
QY 200 GSPISQKIYKENERQYKCNMGYEYSERGDVCTESG-----WRPLPSCEEKSC----- 249
DB 219 GNVPGAIREYKENDVLHYECDRAFKHIDR--PSTCIKQIGKAWSPTPLCESIKCRLTIMD 277
QY 250 -----DNPIYP-----NGDYS----- 260
DB 278 GTRYEPAYRNLFSPGETLVKICARTSWISTPQTSVVTTCQDNGEWSIRPTQOEVRCSNR 337
QY 261 -PLRI-----KHTGDEITYOCNRYFPATRGNT-AKCTSTGWIPAPRCTLRKPCDY 309
DB 338 RPEHVDSDVRSWERYTLDNTRYCKRG-YKRTGGVWATCGRNGMNPCLCEVTKCSK 396
QY 310 PDIKHG-----GLYHENWRRPY 326
DB 397 ENIQDAVIVGTDKQIYNLNOKAIY 420

RESULT 5
Q28769 PRELIMINARY; PRT; 522 AA.
ID Q28769
AC Q28769
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE COMPLEMENT RECEPTOR (FRAGMENT).
OS Papio cynocephalus (Yellow baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Papio.
OX NCBI_TaxID=9556;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RA Birmingham D.J., Logar C.M., Shen X.-P., Chen W.;
RT "The baboon erythrocyte complement receptor is a glycoposphatidyl
RT inositol-linked protein encoded by a homologue of the human CRI-like
RT genetic element."
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L77977; AAA99004.1; -.
DR HSSP; P10998; IYVD.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 7.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 7.
KW Receptor.
FT NON_TER
SQ SEQUENCE 522 AA; 56626 MW; 312FCBE03ADF19DC CRC64;
```

```
Query Match      20.1%; Score 377; DB 6; Length 522;
Best Local Similarity 30.2%; Pred. No. 1e-26;
Matches 107; Conservative 55; Mismatches 142; Indels 50; Gaps 22;

QY 3 CN---ELPPRRNTEILTGSWSDQTYPEGTQAIYKCRGYSRLGNVIMVCRKGEWVALNPL 59
DB 48 CNAPEQLPFARPTNLTDAS-----EPFVGYLYKECLPGYHGKPFIIICLKNVWTSARD- 102
QY 60 RKQKRCPCGHPGDPTEGTFLTGNNVFEYGVKAVYTCNEGQYLLGEINRECDTGDG--- 115
DB 103 -KTRKSCRNPDPVNGVWHVI--KDIQFGSQINYSCKNGYRLIGS--SSATCIISGNTVI 158
QY 116 WTNDIPICEYVKC-LPVTAPENKIVSSAMEPDREYH-FGQAVRFVCSNG-----YKIE 167
DB 159 WDNETPICEIIPCGLPPTI-ANGDFIST-----REYFPYGVVYTYRCNLGSRKKLFELV 213
QY 168 GDEMHCS--DD--GFWSKEKPKC-VEISCKSPDVINGSPIS-QKIYKENERFOYKCNM 221
DB 214 GEPSICTSKDDQVGWISGPAQCIIPNKCMPNVENGVLVSVNRSLSLNEVVEFRCP 273
QY 222 GYEYSERGDVCTE-SGWRP-LPSCEEKSCDNPYPNGDYSPLRKHRTGDEITYOCNRY 279
DB 274 GFYMKGPRHVOCALNKPPELPSCSRVCQPPPEILLGEHTPSHQDFSPGOEVFYSCPG 333
QY 280 FYPATRGNTAKCTSTG-WIP-APRCTLKPCD--YPODIKHGGLYHENNRMRYPVPV 329
DB 334 -YDLRGAASLHCTPQGDWNPPEAPCTIVKSCDDELGLQLPHG-----RVLEPL 378

RESULT 6
Q28577 PRELIMINARY; PRT; 3567 AA.
ID Q28577
AC Q28577
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE POLYDOM PROTEIN PRECURSOR.
GN POLYDOM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HENSIC;
RX MEDLINE=20517255; PubMed=11062057;
RA Gilges D., Vinit M.-A., Callebaut I., Coulombel L., Cacheux V.,
RA Romeo P.-H., Vigon I.;
RT "Polydom : a secreted protein with pentraxin, complement control
RT protein, epidermal growth factor and von willebrand factor A
RT domains."
RL Biochem. J. 352:49-59(2000).
DR EMBL; AF206329; AAG32160.1; -.
DR HSSP; P00740; IEDM.
DR MGD; MGI:1928849; Polydom.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR003410; HYR.
DR InterPro; IPR001759; Pentaxin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF02494; HYR; 2.
DR Pfam; PF00084; sushi; 33.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR000895; PENTAXIN.
DR PRINTS; PR00453; VWFADOMAIN.
```



```
DR Pfam: PF00084; sushi; 29.
DR SMART: SM00032; CCP; 29.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_4.
DR PROSITE: PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW Receptor.
FT NON_TER 1911 1911
SQ SEQUENCE 1911 AA; 210174 MW; 535A4DD0EAF521D CRC64;

Query Match 18.7%; Score 351; DB 6; Length 1911;
Best Local Similarity 29.0%; Pred. No. 1.4e-23;
Matches 93; Conservative 57; Mismatches 135; Indels 36; Gaps 21;

QY 20 SDQTYPEGTQAIYKCRGYSRSLGNIMV-CRKG-EKVALNPLRKCQKRGHPGDTPEGT 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1299 NDFEFPVETSLNYECPGV--PGRMFSISCLNVLWSSVED--NCRKSKCGTPPE-PFNG 1353

QY 78 FTLTGGNVFEYGVKAVYTCNCGYQLLGEINRECDTGG----WTNDIPICEVVKLPVTA 133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1354 MYHINTDT-OFGSTVNYSCNCGFRILGSPS--TTCLVSGNNVTWDEAPICEIISCKPPPT 1411

QY 134 PENGKIVSSAMPEPDREYHFGQAVRFVCSNG-----YKIEGDEEMHCS--DD--GFWSKE 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1412 ISNGDFYSNNR--TSFHSQVTVTYQHTGPDGEQLFELVGERSIYCTSKDQVGNWSSP 1468

QY 184 KPKCVKELIS-CKSPDVINGSP1-SQKIIYKENERFOYKCNMGVEYSEKDAVC-TESGNRP 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1469 PPRCISTNKTAPEKIRVPGNRSFTSLNIVRFPQPGFVVGSHTSYVQCQTNNRWP 1528

QY 241 -LPSCEEKSCDNPYPNGDYSPL-RIKHRTGDEITYQCRNGFYPATRGNTAKCTSTG-WI 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1529 KLPHCSRVCOPPELHGEHTPSHDKFSPQGEVYFSCPEG-YDLRGAASLHCTPGQDWS 1587

QY 298 P-APRCTLKPCD--YPDTKHG 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1588 PEAPICTVKSCDFELGQLPHG 1608

RESULT 9
QSUQV2 PRELIMINARY; PRT; 559 AA.
ID Q9UQV2 AC Q9UQV2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CR1 C3B/C4B RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=69010527; PubMed=2971757;
RA Hourcade D., Miesner D.R., Atkinson J.P., Holers V.M.;
RT "Identification of an alternative polyadenylation site in the human
RT C3b/C4b receptor (complement receptor type 1) transcriptional unit and
RT prediction of a secreted form of complement receptor type 1.";
RL J. Exp. Med. 168:1255-1270(1988).
DR EMBL; X14362; CAA32541.1; JOINED.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF00084; sushi; 8.
DR PRINTS; PRO0343; SELECTIN.
DR SMART; SM00032; CCP; 8.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
KW Signal; Receptor.
FT NON_TER 1 16
FT SIGNAL <1 16 POTENTIAL.
FT CHAIN 17 559 POTENTIAL.
FT SEQUENCE 559 AA; 61424 MW; DBFFE965CA179D75 CRC64;

Query Match 18.7%; Score 350.5; DB 4; Length 559;
Best Local Similarity 29.3%; Pred. No. 3.3e-24;
Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;

QY 3 CNE---LPPRRNTEILTGSWSQTYPEGTQAIYKCRGYSRSLGNIMVCRKGEWVALNPL 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 CNAPEWLFPARPTNLT---DEFEFPDGYLNYECPGYSGRPFSTICLNKSNVMTGAKD- 72

QY 60 RKCQRPCGHPGDTPEGTFTLTGGNVFEYGVKAVYTCNCGYQLLGEINRECDTGG----DTDG 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 73 -RCRRKSCRNPDPVNGMVHVIKNG--IQGSIKYSCTKGYRLIGS--SSATCIISGDTVI 128

QY 116 WTNDIPICSVKRC-LPVTAPENGKIVSSAMPEPDREYHFGQAVRFVCSNG-----YKIEG 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 WNEPTICDRIFCGLPPTI-TGNDFISTNRE---NFHYGSVVTYRCNPGSGGRKVFELVG 184

QY 169 DEMHC--SDD--GFWSKEKPKC-VEISKSPDVINGSPIS-OKIIYKENERFOYKCNMG 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 185 EPSIYCTSDNDQGVIGWSGPAQCIIIPNCKTPPNVENSILVSDNRSLSFSLNEVVEFCQPG 244

QY 223 VEYSERGDVACTE-SQWRP-LPSCEEKSCDNPYIPN-----GDYSPLRKIHRTG 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 FVWKGPRRVKCOALNKWPELPSC-SRVCQPP--PDVLAERTQRDKDNFSP-----G 294

QY 270 DEITYQCRNGFYPATRGNTAKCTSTG-WIP-APRCTLKPCD 308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 QEVFYSCPEG-YDLRGAASMRCTPGQDWSPPAAPTCEVKSCD 334

RESULT 10
Q16745 PRELIMINARY; PRT; 2039 AA.
ID Q16745 AC Q16745;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE COMPLEMENT RECEPTOR 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94065175; PubMed=8245463;
RA Vik D.P., Wong W.W.;
RT "Structure of the gene for the F allele of complement receptor type 1
RT and sequence of the coding region unique to the S allele.";
RL J. Immunol. 151:6214-6224(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Vik D.P., Wong W.W.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L17418; AAB60694.1; JOINED.
DR EMBL; L17390; AAB60694.1; JOINED.
DR EMBL; L17399; AAB60694.1; JOINED.
DR EMBL; L17409; AAB60694.1; JOINED.
DR EMBL; L17419; AAB60694.1; JOINED.
DR EMBL; L17420; AAB60694.1; JOINED.
DR EMBL; L17421; AAB60694.1; JOINED.
DR EMBL; L17422; AAB60694.1; JOINED.
DR EMBL; L17423; AAB60694.1; JOINED.
DR EMBL; L17391; AAB60694.1; JOINED.
DR EMBL; L17392; AAB60694.1; JOINED.
DR EMBL; L17393; AAB60694.1; JOINED.
DR EMBL; L17394; AAB60694.1; JOINED.
DR EMBL; L17395; AAB60694.1; JOINED.
DR EMBL; L17396; AAB60694.1; JOINED.
DR EMBL; L17397; AAB60694.1; JOINED.
DR EMBL; L17398; AAB60694.1; JOINED.
DR EMBL; L17400; AAB60694.1; JOINED.
DR EMBL; L17401; AAB60694.1; JOINED.
DR EMBL; L17402; AAB60694.1; JOINED.
```



```
QY 169 DEEMHC--SDD--GFWSKPKKC-VEISCKSPDVINGSPIS-OKIYKENERFOYKCNMG 222
Db 210 ESIYCTSDNDQGVIGSWGPAPOCIIPNKCTPPNVENGILVSDNRSLFSLNVEVFRCPQC 269
QY 223 YEYSEGDVACTE-SGWRP-LPSCCEKSCDNPYIIN-----GDYSLRLKHRGTG 269
Db 270 FVMKGPVRVKCOALNKWPELPSC-SRVCOPP--PDVLHAERTQDKDNFSP-----G 319
QY 270 DEITYQCRNGFYPATRGNTAKTSTG-WIP-APRCTLKPCD 308
Db 320 QEVFYSCEPG-YDLRGAASMRCTPGQDWSPPAAPTCEVRKSD 359

RESULT 12
Q9WRU2
ID Q9WRU2 PRELIMINARY; PRT: 645 AA.
AC Q9WRU2;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE COMPLEMENT BINDING PROTEIN.
OS Macaca mulatta rhadinovirus 17577.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=83534;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99174001; PubMed=10074154;
RA Searles R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF083501; AAD21332.1;
DR HSSP: P10998; 1VDV
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 8.
DR SMART: SM00032; CCP; 8.
SQ SEQUENCE 645 AA; 71526 MW; 93D8DE35ABF61EB2 CRC64;

Query Match 18.6%; Score 349.5; DB 12; Length 645;
Best Local Similarity 25.8%; Pred. No. 4.9e-24;
Matches 99; Conservative 45; Mismatches 164; Indels 75; Gaps 14;

QY 1 EDCNELPPRRNTEILTGSMSDQTYPEGTQAIYKCRPGYRSLGNVIMV--CRKGEWALNP 58
Db 21 ENCK--PPHFTYRKYKSTKDLISVGTEALICRPGYVNTKIITTECLONGTW--STP 76
QY 59 LRKQKRCPCGHPGDTFTLTG--NVFEYGVKAVYTCNEGQYLLGEINREC---DTD 114
Db 77 NFPCDKRCPTPADLLNGAVIHGGDNALKFGSNISYECNEGDLIGS--NVRFCLQDTE 135
QY 115 --GWTNDIPICEVVKLPVTAPENKIVSSAMEPDRE-YHFGQAVRVCNSGYKIEGDEE 171
Db 136 NVNWDNSPEVCEIOKICRPPAVEHGDVL-----PNQDVYNGDAITFKCSLSTLVGSTT 190
QY 172 MHCSDGDFWSEKPKFCVKEISCKSPDVING-SPISOKIYKENERFOYKCNMGYSEKGD 230
Db 191 LVCTSNKWNNSFTCLMLVCESPQIDNGYIDIGLSRRYNHGQSITVKCSGYNIVGPET 250
QY 231 AVCTESGW-RPLPSC-----EEKSCDNPYIPNGDYSPLRI----- 264
Db 251 LTCNTTWTVPPLPKCVLVTNNPSTPMPETPTPDYQKINLSTAKTATTPNAFTVT 310
QY 265 -----KRTGDEITYQCRNGFYPATRGNTAKTSTGWI 297
Db 311 VSPEKDDVTCVPHFRFMVKAENDKEKYSVGASVELICRPGFTKMQSTVSVECLSNGTW 370
```

```
QY 298 PAP--RCTLKPCDYPDIDIKHGGLY 318
Db 371 TAPNAKCHRRKCKPTQOELLNGEY 393

RESULT 13
Q29530
ID Q29530 PRELIMINARY; PRT: 2014 AA.
AC Q29530;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94292799; PubMed=8021505;
RA Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;
RT "Primary sequence of an alternatively spliced form of CRI. Candidate
RT for the 75,000 M(r) complement receptor expressed on chimpanzee
RT erythrocytes."
RL J. Immunol. 153:691-700(1994).
DR EMBL: L24920; AAA51438.1;
DR HSSP: P08603; 1HEI
DR InterPro: IPR001424; SOD_CU_ZN.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR InterPro: IPR000834; Zn_carboOpept.
DR Pfam: PF00084; sushi; 30.
DR SMART: SM00032; CCP; 30.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE: PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW Receptor.
FT NON_TER 1 2014
FT NON_TER 2014 2014
SQ SEQUENCE 2014 AA; 221281 MW; 6D6C3A74D81F1DB9 CRC64;

Query Match 18.6%; Score 348; DB 6; Length 2014;
Best Local Similarity 28.4%; Pred. No. 2.7e-23;
Matches 96; Conservative 57; Mismatches 135; Indels 50; Gaps 23;

QY 20 SDQTYPEGTQAIYKCRPGYRSLGNVIMV--CRKG-EMVALNPLRKCKOKRCPGHGDTPTFGT 77
Db 1387 NDFEPVGTSLNYECPGT--FGKMFISCLLENLVSSVED--NCRKSCGPPPE-PFNG 1441
QY 78 FTLTGNNVFEYGVKAVYTCNEGQYLLGEINRECDDTG----WTNDIPICEVVKCLPVTA 133
Db 1442 MVHINTDT-QFGSTVYSCNEGFRILIGSPS-TTCLVSGNNVTMDKKAPICEIICEPPT 1499
QY 134 PENGKIVSSAMEPDREYHFGQAVRVCNSG-----YKIEGDEMHCS--DD--GFWSKE 183
Db 1500 ISNGDFYSNNR---ASFHNGTVVTYQCHTGPDEGLFELVGRSICYTSKDDQGVWVSSP 1556
QY 184 KPKCVEIS-CKSPDVINGSPI-SOKIYKENERFOYKCNMGYSEKGDVACTESG-WRP 240
Db 1557 PPRCISTNKCTAPEVNAIRVPGNRSFSLTEIVRFRCPGFMVGSHTVQCOTGRMGP 1616
QY 241 -LPSCCEKSCDNPYIPNGDYSPLRIKHR-----TGDEITYQCRNGFYPATRGNTAKTSTG 295
Db 1617 KLPHCSRVCCPPPELHGEHTP---SHQDNFSPQGEVYFSCPEG-YDLRGAASLHCTPQG 1672
QY 296 -WIP-APRCTLKPCD--YPDIKHGGLYHENMRRPYFPV 329
Db 1673 DWSPEAPRCTVKSCDDFLGQLPHG-----RVLFPL 1702

RESULT 14
Q29531
```

Q29531 PRELIMINARY; PRT; 661 AA.

Q29531; AC

Q29531; DT

01-NOV-1996 (TrEMBLrel. 01, Created)

01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE COMPLEMENT RECEPTOR 1 PRECURSOR (FRAGMENT).

GN

OS Pan troglodytes (Chimpanzee).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

OX NCBI_TaxID=9598;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=94292799; PubMed=8021505;

RA Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;

RT "Primary sequence of an alternatively spliced form of CRI. Candidate

RT for the 75,000 M(r) complement receptor expressed on chimpanzee

RT erythrocytes."

RL J. Immunol. 153:691-700(1994).

DR EMBL; L24921; AAA51439.1; -

DR HSSP; P10998; IYVD.

DR InterPro; IPR002396; Selectin.

DR InterPro; IPR001424; SOD_CU_ZN.

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam; PF00084; sushi; 9.

DR PRINTS; PR00343; SELECTIN.

DR SMART; SM00032; CCP; 9.

DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.

KW Signal; Receptor.

FT NON_TER 1

FT SIGNAL <1

FT CHAIN 17

FT CHAIN 661

FT NON_TER 661

FT SEQUENCE 661 AA; 72966 MW; 3F0DF1532664D3F0 CRC64;

Query Match 18.3%; Score 342.5; DB 6; Length 661;

Best Local Similarity 29.0%; Pred. No. 2.2e-23;

Matches 99; Conservative 55; Mismatches 128; Indels 59; Gaps 23;

QY 3 CNE---LPPRNTTELTGSDQYVPECTQAIYKCRPGYRSLGNVIMVCRKEWVALNPL 59

Db 18 CNAEWLPFAPRTNLT-----DEFEPITGLYNECRPGYGRPFISILCKNSVWFGAKD- 72

QY 60 RKQKRCPGHPGDPFGFTTGTGTVNFYGVKAVYTCNEGYSQLLGEINYNREC-----DTDG 115

Db 73 -RCRRKSCRNPDPVNGMVHVI--KDIOFGSQIKYCTKGYRLIGS--SSATCIISGDVVI 128

QY 116 WTNDIPICEVVKC-LPVTAPENGKIVTSSAMPDREYHFGQAVRVCNSG-----YKIEG 168

Db 129 WDNETPICDRIPICGLPPTI-TNGDFISNRE---NFHGSVVTYRCNPGSGGRKVFELVG 184

QY 169 DEEMHC--SDD--GFWSKEKPC-VEISCKSPDVINGSPIS-OKIIYKENERFQYKCNNG 222

Db 185 EPSIVCTSNDDQVIGWSGAPQCIIIPNKCTPPNVENGILVSDNRSLSFNEVVEFRCPG 244

QY 223 YEYSERGAUCTE-SGWRP-LPSCSEKSCDNPYPN-----GDYSPLRIKHTG 269

Db 245 FVMKGPVRVKQALNKWEPELPSC-SRVQCPP--PDVLHAERTQDKDNFSP-----G 294

QY 270 DEITYQCRNGFYPATRGNTAKTSTG-WIPA-PRCTLKPCD 308

Db 295 QEVFYSCEPG-YDLRGAASLCTPGDWSPATPTCEVAKSCD 334

RESULT 15

Q28797 PRELIMINARY; PRT; 522 AA.

Q28797; AC

Q28797; DT

01-NOV-1996 (TrEMBLrel. 01, Created)

01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE MRNA SEQUENCE, 3' END OF ORF (FRAGMENT).

DE

OS Pan troglodytes (Chimpanzee).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

OX NCBI_TaxID=9598;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=94292799; PubMed=8021505;

RA Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;

RT "Primary sequence of an alternatively spliced form of CRI. Candidate

RT for the 75,000 M(r) complement receptor expressed on chimpanzee

RT erythrocytes."

RL J. Immunol. 153:691-700(1994).

DR EMBL; L24922; AAS50460.1; -

DR HSSP; P10998; IYVD.

DR InterPro; IPR002396; Selectin.

DR InterPro; IPR001424; SOD_CU_ZN.

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam; PF00084; sushi; 9.

DR PRINTS; PR00343; SELECTIN.

DR SMART; SM00032; CCP; 9.

DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.

KW Signal; Receptor.

FT NON_TER 1

FT SIGNAL <1

FT CHAIN 17

FT CHAIN 661

FT NON_TER 661

FT SEQUENCE 661 AA; 72966 MW; 3F0DF1532664D3F0 CRC64;

Query Match 18.1%; Score 339; DB 6; Length 522;

Best Local Similarity 30.1%; Pred. No. 3.5e-23;

Matches 95; Conservative 52; Mismatches 119; Indels 50; Gaps

QY 24 YPEGTOAIYKCRPGYRSLGNVIMVCRKEWVALNPLRCKOKRCPGHPGDPFGFTTGTG 83

Db 5 FPIGTYLNECRPGYGRPFISILCKNSVWTSAD--KCKRKSERNPPDPVNGMVHVI-- 60

QY 84 NVFEGVKAVYTCNEGYSQLLGEINYNRECDDTG-----WTNDIPICEVVKC-LPVTAPENGK 138

Db 61 KDIOFGSQIKYCTKGYRLIGS--SSATCIISGNTVIWDNKTVCYDRIICGLPPTI-ANGD 118

QY 139 IVSSAMEPDREY-HFGQAVRVCNSG-----YKIEGDEMHCS--DD--GFWSKEKPC 187

Db 119 FTSIS-----REFHYASVVTYHCNLGSGGKVFELVGEPSIYCTSKDDQVIGWSGAPQC 174

QY 188 -VEISCKSPDVINGSPIS-OKIIYKENERFQYKCNMGYYSERGAUCTE-SGWRP-LPS 243

Db 175 IIPNKCTPPNVENGILVSDNRSLSFNEVVEFRCPGFMKGPVRVKQALNKWEPELPS 234

QY 244 CEKSCDNPYPNGD-----YSPRLRIKHTGDEITYQCRNGFYPATRGNT-AKCTST 294

Db 235 CSRVQCPPDVLHGERTQDKDNFSP-----GEEVYSCEPGY--DLRGSTLYLHCTQP 285

QY 295 G-WIP-APRCTLKPCD 308

Db 286 GDWSPEAPRCEVAKSCD 301

Search completed: August 29, 2002, 15:11:41

Job time: 602 sec

Search completed: August 29, 2002, 15:11:41
Job time: 602 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:03:37 ; Search time 75.55 Seconds
(without alignments)
483.697 Million cell updates/sec

Title: US-09-316-163-11

Perfect score: 1876

Sequence: 1 EDCNELPPRRNTTEILTGWS.....PDIKHGGLYHENRRPYFPV 329

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1362	72.6	240	AAW39154	Human partial Comp
2	1229	65.5	216	AAW39155	Clone PRR9FH410 C
3	353	18.8	496	AAW55752	Human CR1 protein
4	353	18.8	581	AAAI3490	Human C4 binding p
5	352.5	18.8	543	AAW28570	CR1-4 (266-274 KLK
6	350.5	18.7	778	AAW73147	Amino acid sequenc
7	350.5	18.7	1930	AAW45899	Human complement r
8	350.5	18.7	2039	AAW55751	Human C3b/C4b rece
9	350.5	18.7	2039	ABG00287	Novel human diagno
10	350.5	18.7	2044	ABB11782	Human CR1 protein
11	350.5	18.7	2044	AAW39224	Human polypeptide

12	350.5	18.7	2044	22	AAW41010	Human polypeptide
13	350.5	18.7	2317	10	AAW92219	CR1 protein. Homo
14	349.5	18.6	645	21	AAW53125	Macaca mulatta rha
15	347.5	18.5	543	13	AAW28557	CR1-4 (99H, 103E)
16	345.5	18.4	543	13	AAW28547	CR1-4 (52S, 53S, 5
17	343.5	18.3	543	13	AAW28560	CR1-4 (114S) analo
18	341.5	18.2	543	13	AAW28550	CR1-4 (64K) analog
19	341.5	18.2	543	13	AAW28553	CR1-4 (85R, 87N) a
20	341.5	18.2	543	13	AAW28565	CR1-4 (121Q) analo
21	341.5	18.2	543	13	AAW28568	CR1-4 (347T, 349Y)
22	341.5	18.2	543	13	AAW28569	CR1-4 (369-376 STK
23	341.5	18.2	543	13	AAW28571	CR1-4 (364-367 NAA
24	341.5	18.2	2039	12	AAW11810	Human complement t
25	340.5	18.2	543	13	AAW28545	CR1-4 (37Y) analog
26	340.5	18.2	543	13	AAW28548	CR1-4 (57V, 59K) a
27	340.5	18.2	1497	22	AAW93954	Human polypeptide,
28	339.5	18.1	543	13	AAW28555	CR1-4 (92T) analog
29	339.5	18.1	543	13	AAW28558	CR1-4 (109N, 110A,
30	339.5	18.1	543	13	AAW28563	CR1-4 (117P) analo
31	339.5	18.1	543	13	AAW28567	CR1-4 (318-321 RNP
32	338.5	18.0	543	13	AAW28549	CR1-4 (64K, 65T) a
33	338.5	18.0	543	13	AAW28551	CR1-4 (65T) analog
34	338.5	18.0	543	13	AAW28562	CR1-4 (116K) analo
35	336.5	17.9	543	13	AAW28556	CR1-4 (94H) analog
36	336.5	17.9	543	13	AAW28564	CR1-4 (116K, 117P)
37	336.5	17.9	2039	14	AAW36743	CR1. Homo sapiens
38	336	17.9	1537	12	AAW11982	Partial human comp
39	336	17.9	1929	22	ABG00103	Novel human diagno
40	335.5	17.9	543	13	AAW28552	CR1-4 (78T, 79D) a
41	335.5	17.9	543	13	AAW28561	CR1-4 (115T) analo
42	335.5	17.9	543	13	AAW28566	CR1-4 (318R, 319N)
43	334.5	17.8	543	13	AAW28554	CR1-4 (92T, 94H) a
44	333.5	17.8	543	13	AAW28544	CR1-4 (35E) analog
45	333.5	17.8	543	13	AAW28546	CR1-4 (44T, 47D, 4

ALIGNMENTS

RESULT 1
AAW39154
ID AAW39154 standard; Protein; 240 AA.
XX
AC AAW39154;
XX
DT 27-APR-1998 (first entry)
XX
DE Human partial Complement factor H protein fragment 1.
XX
DE Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator.
XX
OS Homo sapiens.
XX
PN WO9730136-A1.
XX
PD 16-OCT-1997.
XX
PF 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
PR 09-APR-1996; 96US-0015083.
PR 09-APR-1996; 96US-0630048.
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
PI Enfield DL, Hass GM, Kinders RJ;
XX
DR WPI; 1997-512742/47.
DR N-PSDB; AAV02790.
XX
PT Treating or screening for cancer, e.g. renal or urogenital cancer -

PT by modulating or detecting tumour associated human complement Factor
PT H related antigen, or nucleic acid encoding it
XX
PS Example 6B; Fig 6B; 104pp; English.
XX
CC This partial protein sequence represents a region of the human
CC tumour-associated complement factor H (CFH). This sequence is used
CC in the identification of complement factor H related proteins and
CC antigens isolated from clone PRB9FH410 (see AAW39155). The detection of
CC such proteins and a CFH antigens can be used in screening or for the
CC treatment of renal or urogenital cancer, e.g. bladder, cervical or
CC prostate cancer. Agents that may modulate this antigen could be used in
CC the manufacture of a medicament for the treatment of a tumour cell.
XX
SQ Sequence 240 AA;

Query Match 72.6%; Score 1362; DB 18; Length 240;
Best Local Similarity 100.0%; Pred. No. 3.7e-97;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 FTLTGNVFEYGVKAVYTCNEGYYQLLGEINRECDTGTNDIPICEVVKCLPVTAPENG 137
|||||
Db 1 fultgnvfeygvkavycnecyqllegeinyrecdtgtndipicevvkclpvtapeng 60
QY 138 KIVSSAMEPDRYHFGQAVRFVNCNSGYKIEGDEEMHCSDDGFWSEKPKCVCISCKSPDV 197
|||||
Db 61 kivssamepdrayhfgqavrfvcnsgykiegdeemhcsddgfwsekepkcveiscspdv 120
QY 198 INGSPIOKIYKENERFOYKCNMGYEXSERGDVACTESGWRPLPSCSEKSCDNPYIPNG 257
|||||
Db 121 ingspisokliykeneryfocnmgeyesergdgvactesgwrplpsceekscdnpyipng 180
QY 258 DYSPLRIKHTGDEITYQCRNGFYPATRGNTAKCTSTGWIAPAPRCTLPCDYPDIKHGGL 317
|||||
Db 181 dysplrikhtgdeityqcrngfyatrgntaktstgwiapaprcctlkpcdydikhagl 240

RESULT 2
AAW39155
ID AAW39155 standard; Protein: 216 AA.
XX
AC AAW39155;
XX
DT 27-APR-1998 (first entry)
XX
DE Clone PRB9FH410 CFH related protein fragment.
XX
KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator.
XX
OS Synthetic.
XX
PN WO9738136-A1.
XX
PD 16-OCT-1997.
XX
PF 09-APR-1997; 97WO-US05710.
XX
PR 06-MAR-1997; 97US-0812481.
PR 09-APR-1996; 96US-0015083.
PR 09-APR-1996; 96US-0630048.
PR 06-MAR-1997; 97US-0038614.
XX
PA (BARD-) BARD DIAGNOSTIC SCI INC.
XX
PI Enfield DL, Hass GM, Kinders RJ;
XX
DR WPI: 1997-512742/47.
DR N-PSDB; AAV02791.
XX
PT Treating or screening for cancer, e.g. renal or urogenital cancer -
PT by modulating or detecting tumour associated human complement Factor

PT H related antigen, or nucleic acid encoding it
XX
PS Example 6B; Fig 6B; 104pp; English.
XX
CC This partial protein is found in clone PRB9FH410 and represents a
CC complement factor H related protein with homology to a region of the
CC human tumour-associated complement factor H (CFH). The detection of this
CC protein and a CFH antigen can be used in screening or for the treatment
CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
CC Agents that may modulate this antigen could be used in the manufacture of
CC a medicament for the treatment of a tumour cell.
XX
SQ Sequence 216 AA;

Query Match 65.5%; Score 1229; DB 18; Length 216;
Best Local Similarity 99.5%; Pred. No. 5.7e-87;
Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 95 TCNEGYYQLLGEINRECDTGTNDIPICEVVKCLPVTAPENGKIVSSAMEPDRYHFGQ 154
|||||
Db 1 tcnegyqllegeinyrecdtgtndipicevvkclpvtapengkivssamepdrayhfgq 60
QY 155 AVRFVNCNSGYKIEGDEEMHCSDDGFWSEKPKCVCISCKSPDVINGSPISQKIIYKENER 214
|||||
Db 61 avrfvcnsgykiegdeemhcsddgfwsekepkcveiscspdvingspsqkiiykener 120
QY 215 FOYKCNMGYEXSERGDVACTESGWRPLPSCSEKSCDNPYIPNGDYSPLRIKHTGDEITY 274
|||||
Db 121 foykcnmgeyesergdgvactesgwrplpsceekscdnpyipngdysplrikhtgdeity 180
QY 275 QCRNGFYPATRGNTAKCTSTGWIAPAPRCTLPCDYP 310
|||||
Db 181 qcrngfyatrgntaktstgwiapaprcctlkpcdyp 216

RESULT 3
AAV55752
ID AAV55752 standard; Protein: 496 AA.
XX
AC AAV55752;
XX
DT 22-FEB-2000 (first entry)
XX
DE Human CR1 protein LHR-A SCR fragment.
XX
KW C3B/C4B receptor; CR1 protein; cell-surface protein; erythrocyte; human;
KW complement regulatory activity; complement pathway enzyme; tissue damage;
KW reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;
XX
OS Homo sapiens.
XX
PN US5981481-A.
XX
PD 09-NOV-1999.
XX
PF 06-JUN-1995; 95US-0470652.
XX
PR 03-APR-1989; 89US-0332865.
PR 06-DEC-1974; 74US-0350238.
PR 24-FEB-1993; 93US-0026134.
PR 01-APR-1988; 88US-0176532.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI Concino MF, Wong WM, Makrides SC, Klickstein LB, Fearon DT, Ip SH;
PI Marsh HC, Carson GR;
XX
DR WPI: 1999-633357/54.

WPI; 1998-568350/48.

Fragment of soluble human complement receptor 1 - useful for treating T-cell or B-cell mediated immune responses e.g. inflammatory responses such as rheumatoid arthritis

Disclosure; Fig 1; 54pp; English.

This is an amino acid sequence of the human soluble complement receptor 1 (sCR1), useful in the treatment of T-cell or B-cell mediated immune responses. It is used to inhibit a T-cell or B-cell-mediated immune response or prevent immune response-mediated tissue rejection and destruction or clearance or inactivation of an expressed protein especially from cells that have been treated by gene therapy to express the protein, e.g. dystrophin. The protein can also be used to inhibit a T-cell or B-cell-mediated inflammatory response, an interferon-gamma secretory response, autoimmune response or neurological response, e.g. Alzheimer's or Parkinson's disease or multiple sclerosis. Also the protein can be used to treat systemic lupus erythematosus, rheumatoid arthritis, myasthenia gravis, epidermis bullosa or Hashimoto's disease.

Sequence 778 AA:

[illegible]

RESULT	7	
AAW45899		
ID	AAW45899 standard; peptide; 1930 AA.	
XX		
XX		
XX	AAW45899;	
XX		
DT	30-JUN-1998 (first entry)	
DE	Human complement receptor 1 (residues 1-1929).	
XX		
XX		
DE		
XX		
KW	Membrane binding element; thrombotic disease; soluble protein;	
KW	complement-related disease; integral membrane protein; inflammation;	
KW	short consensus repeat; SCR 1-3; CR1; complement receptor type 1.	
XX		
OS	Homo sapiens.	

	Key	Location/Qualifiers 1930
xx		
FH		
FT	Cross-links	

FT	/note= "Disulphide linked to Cys in peptide given
FT	in AAW45889"
PN	
PN	WO9802454-A2.
XX	
PD	22-JAN-1998.
XX	
XX	08-JUL-1997; 97WO-EP03715.
XX	
XX	15-JUL-1996; 96GB-0014871.
XX	
PA	(ADPR-) ADPROTECH PLC.
XX	
XX	Dodd I, Mossakowska DEI, Smith RAG;
XX	
XX	WPI; 1998-110524/10.
XX	
XX	Derivatives of soluble poly:peptide(s) bonded to low affinity
PT	membrane binding groups - useful for treating complement-related and
PT	thrombotic diseases, providing improved localisation at cellular
PT	membranes
XX	
PS	Claim 22; Pages 60-61; 75pp; English.
XX	
CC	This sequence represents human complement receptor 1 (CR1, CD 35)
CC	N-terminal fragment. The invention relates to a soluble derivative (A)
CC	of a soluble polypeptide (I), which comprises at least 2 heterologous
CC	membrane-binding elements (MBE) of low membrane affinity covalently
CC	associated with (I). MBE interact, independently and with thermodynamic
CC	additivity, with components of cellular or artificial membranes exposed
CC	to extracellular fluids. (A) are used to treat disorders treatable with
CC	(I) itself, specifically inflammation or any other complement-related
CC	disorder (e.g. neurological disease, graft rejection, myocardial
CC	infarction, sepsis, rheumatoid arthritis and many others; including
CC	application to indwelling devices) and thrombolytic disease, but also to
CC	treat allergy, induce weight loss, to treat ischaemia or asthma and as
CC	immuno-modulators for treating multiple sclerosis. (A) are administered
CC	orally, topically, by injection or inhalation at 0.01-10 (preferably
CC	0.1-10) mg/kg/day.
XX	
SQ	Sequence 1930 AA.
SQ	

```

Query Match      18.7%   Score 350.5;   DB 19;   Length 1930;
Best Local Similarity 29.3%;   Pred. No. 3.3e-18;
Matches 100;   Conservative 55;   Mismatches 127;   Indels 59;   Gaps 23;

QY      3 CNE---LPPRNTETILGTSNSDQTYPECTQAIYKCRPGYSLGNVIMVCRRGEWALNPL 59
       || | | | | : : | | | | | | | | : : |
Db      2 cnapewlpfartnlt----defeflgtlylnyecrpgysgrfsilclknsvwtgakd- 56
       || : : : : : : : : : : : : : : : : : :

QY      60 RKCKRPCHGPDPNPFGTFTLTGNGVFVEYGVKAVVTGNEGYYOLLGEINVRBC-----DFDG 115
       || : : : : : : : : : : : : : : : : : :
Db      57 -rcrrksrnpdpvngmvhvikg--lqfgsqklyscctkyrligs-ssatcilsgdvlv 112
       || : : : : : : : : : : : : : : : : : :

QY      116 WTNDIPICEVKKC-LPVTAPENGKIVSAMBPDRHYHFGQAVRFVCNSG-----YKIEG 168
       || | | | | : : | | | | : : | | | | : : |
Db      113 wdnetspicdlpcglpbpti-tngafistnr--nfnygsvtvrcnpgsggrkvfelvg 168
       || : : : : : : : : : : : : : : : : : :

QY      169 DEEMHC--SDD-GFWSEKPKC-VEISCKSPDVINGSPIS-QKIIVKENERFOYKCNMG 222
       || : : || | | | | : : | | | | : : | | : : |
Db      169 epslytsnddqgiwggppaqcplpnktcpnvengllvsdnrsifslnevfvefrccpq 228
       || : : || | | | | : : | | | | : : | | : : |

QY      223 YEYSERGDVACTE-SGWRP-LPSCEEKSKCONPYLPN-----GDYSPLRIKURTG 269
       || : : || | | | | : : | | | | : : | | : : |
Db      229 fvmkgprvrvcqalnkweelpsc-srvccpp-pdvlhaertqrkdnfsp-----g 278
       || : : || | | | | : : | | | | : : | | : : |

QY      270 DETIYQCRNGFYPATRGNTAKCTSTG-WIP-APRCTLKPCD 308
       || : : || | | | | : : | | | | : : | | : : |
Db      279 qevfyscepq-ydlrqasmrctbpqqdwspaaptcevkscd 318
       || : : || | | | | : : | | | | : : | | : : |

```

RESULT

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 2039 AA;

Query Match	18.7%	Score	350.5	DB	22	Length	2039
Best Local Similarity	29.3%	Pred. No.	3.6e-18				
Matches	100	Conservative	55	Mismatches	127	Indels	59
						Gaps	23

[illegible]

RESULT 10

ABB11782
ID ABB11782 standard; peptide; 2044 AA.

AC ABB11782;

DT 11-JAN-2002 (first entry)

Human CR1 protein homologue, SEQ ID NO:2152.

Human; cytokine; cell proliferation; cell differentiation; growth factor;
haematopoiesis regulation; tissue growth; immunomodulator; activin;
inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
cell proliferation; metastasis; cancer; tumour; haematopoietic disorder;
myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
chronic inflammatory condition; proliferative retinopathy;
atherosclerosis; coronary heart disease; arterial ischaemia;
bone disorder; osteoporosis; vascular growth disorder;
tissue regeneration; wound healing; infection; immune disorder;
cell culture; drug screening; gene therapy; antiinflammatory;
antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;
cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
antifungal; vulnery; antiulcer.

OS Homo sapiens.

PN WO200157188-A2.

PD 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US03800.

PR 03-FEB-2000; 2000US-0496914.

XX
27-AFK-2000; 200003-03000873.XX
 (HIGH) HIGH INC.

PI Tang YT, Liu C, Drmanac RT;

[illegible]

Query Match 18.7%; Score 350.5; DB 22; Length 2044;
Best Local Similarity 29.3%; Pred. No. 3.6e-18;
Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;
QV 3 CNE---LPPRRNTEILGSSNDOTYPEGTOAIYKCRPCYRSIGNVIMVCRKGEVVALNPL 59

```
Qy      3 CNE--LPPRNTEILTGSWSDQTPGQTQAIYKCRPGYRSLGNVIMVCRKGEWALNPL 59
        ||   ||   ||   :   ||   ||   ||   ||   ||   ||   ||   ||   |
Db     48 cnapewlpfartnlt---defeflqtlvncrpgysgrpfslclcnksvwtaakd- 102
```

```

Qy 60 RKQKRPCGHPGDPFGTFTLTGGNVFEYGVKAVYTCNEGYYQLLGEINYREC---DTD 115
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 103 -rarrkscrnppdpvngmvhvkq--lqfsgtkysctkgyrligs-ssatclisgdtvi 158

```

Qy	116	WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYHFQQAVERFCVNSG-----YKIEG	168
		: : : : : : : :	
Db	159	wdnetpicdrpcglppti-tngdfistnre---nfhygsvvttyrcnpgsggrkvvclvg	214

Qy	169	DEEMHC--SDD--GFWSKERPKC-VEISCKSPDIVNGSPIS-QKILYKENRFQVKCNMG	222
Db	215	epslyctsnddqvglwsgpapqclpnkctppnvengllvsdnrlslfsinevvefrcpqg	274

QY 223 YEYSERGDAVCTE-SGWRP-LPSCCEKSCDNPYPN-----GDYSPLEIKHRTG 269

WPI: 2001-457740/49.
N-PSDB: ABA09026.

Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject e.g. arthritis and cancer -

Claim 20; Page 245-246; 1963pp; English.

Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoiesis regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

Sequence 2044 AA;

Query Match 18.7%; Score 350.5; DB 22; Length 2044;
Best Local Similarity 29.3%; Pred. No. 3.6e-18;
Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;
QV 3 CNE---LPPRRNTEILGSSNDOTYPEGTOAIYKCRPCYRSIGNVIMVCRKGEVVALNPL 59

```

Qy 3 CNE--LPPRNTEILTGSWSDQTPGQTQAIYKCRPGYRSLGNVIMVCRKGEWALNPL 59
    || || || || || || || || || || || || || || || || || || || || ||
Db 48 cnapewlpfartnlt---defeflqtlvncvcpqysgrpfslclcnksvwtakd- 102

```

```
Qy      60 RKQKRPCGHPGDPPFGTFTLTGGNPFYGVKAVYTCNEGYYQLLGEINYREC---DTD 115
        :|::||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     103 -rarrkscrnppdpvngmvhwikg--lqfsgtkysctkgryllgs-ssatclisgdtvi 158
```

Qy 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDREYHFQOAVRRCVNSG-----YKIEG 168

Db 159 wdnetspdrpcglppti-tngdfistnre---nfhygsvvttyrcnpgsggrkvvclvg 214

Qy	169 DEEMHC--SDD--GFWSKERPKC-VEISCKSPDIVNGSPIS-QKILYKENRFQVKCNMG	222
Db	215 epslyctsnddqvglwsgpapqclpnkctppnvengllvsdnrlslfsinevvefrcpqg	274

QY 223 YEYSERGDAVCTE-SGWRP-LPSCCEKSCDNPYPN-----GDYSPLEIKHRTG 269

XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 XX Example 2; SEQ ID NO 5941; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 2044 AA;

Query Match 18.7%; Score 350.5; DB 22; Length 2044;
 Best Local Similarity 29.3%; Pred. NO. 3.6e-18;
 Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;
 QY 3 CNE---LPRRRTEILTGSWSQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWALNPL 59
 DB 48 cnapewlpfartnlt----defefpigtlyInyecrpdygrpsfslclcknsvwtgkd- 102
 QY 60 RKCKRQPCGHFGDTPFGTTLTGGNVFYGKAVYTCNEGQYLLGEINRYREC----DTDG 115
 DB 103 -rcrrkscrnppdpvngmvhvikg--iqfgsqikysctkgyrligs-ssatclisgdtvi 158
 QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDRYHFGQAVRFVCSNG-----YKIEG 168
 DB 159 wdnetspdrpcglppti-tngdfistnre---nfhygsvvtyrcnpgsggrkvfelvg 214
 QY 169 DEEMHC--SDD--GFWSKEPKC-VEISCKSPDVINGSPIS-OKIYYKENERFOYKCNMG 222
 DB 215 epslyctsnddqvglwsgpapqciipnktppnvengllvsdnrslfsInevveifrcpg 274
 QY 223 YEYSERGDAVCTE-SGWRP-LPSCEEKSCDNPYIPN-----GDYSPLRIKHRTG 269
 DB 275 fvmkgprrvkcqalnkwepelpsc-srvccqp--pdvlhaertqrkdnfsp-----g 324
 QY 270 DEITYQCRNGFYPATRGNTAKTSTG-WIP-APRCTLKPCD 308
 DB 325 qevfyscepg-ydlrgaasmrctpggdwspaaptcevkscd 364

RESULT 13
 AAP92219
 ID AAP92219 standard; protein; 2317 AA.
 XX
 AC AAP92219;
 XX

22-FEB-1990 (first entry)
 CR1 protein.
 XX
 KW Complement; cofactor.
 XX
 OS Homo sapiens (human).
 XX
 FH Key Location/Qualifiers
 FT Peptide 10..50
 FT /label= signal_peptide
 XX

PN W08909220-A.
 XX
 PD 05-OCT-1989.
 XX
 PF 31-MAR-1989; 89WO-US01358.
 XX
 PR 01-APR-1988; 88US-0176532.
 XX
 PA (TCEL) T CELL SCIENCES INC.
 PA (UJO) THE JOHNS HOPKINS UNIVERSITY.
 PA (BRIG) THE BRIGHAM AND WOMEN'S HOSPITAL.
 XX
 PI Fearon DT, Klickstein LB, Wong W, Carson G, Concino WF, Makrides SC;
 XX WPI; 1989-309498/42.
 DR N-PSDB; AAN91477.
 XX
 PT New nucleic acid sequences encoding new CR1 protein - and its fragment,
 PT for diagnosis and control of complement-related immune defects,
 PT inflammation, myocardial infarct, etc
 XX
 PS Claim 1; Fig 1; 191pp; English.
 XX
 CC This is full-length CR1 protein, and shortened forms are new, lacking
 CC the transmembrane region. The proteins and fragments bind C3b and/or
 CC C4b, have cofactor activity and inhibit C3 and C5 convertase activity.
 CC In the sequence, x=untranslated region. This has 7 short consensus
 CC repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in
 CC patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C
 CC for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.
 CC They are useful in diagnosing and treating immune disorders, and prevent
 CC perfusion injury.
 XX
 SQ Sequence 2317 AA;
 Query Match 18.7%; Score 350.5; DB 10; Length 2317;
 Best Local Similarity 29.3%; Pred. NO. 4.1e-18;
 Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;
 QY 3 CNE---LPRRRTEILTGSWSQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWALNPL 59
 DB 52 cnapewlpfartnlt----defefpigtlyInyecrpdygrpsfslclcknsvwtgkd- 106
 QY 60 RKCKRQPCGHFGDTPFGTTLTGGNVFYGKAVYTCNEGQYLLGEINRYREC----DTDG 115
 DB 107 -rcrrkscrnppdpvngmvhvikg--iqfgsqikysctkgyrligs-ssatclisgdtvi 162
 QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDRYHFGQAVRFVCSNG-----YKIEG 168
 DB 163 wdnetspdrpcglppti-tngdfistnre---nfhygsvvtyrcnpgsggrkvfelvg 218
 QY 169 DEEMHC--SDD--GFWSKEPKC-VEISCKSPDVINGSPIS-OKIYYKENERFOYKCNMG 222
 DB 219 epslyctsnddqvglwsgpapqciipnktppnvengllvsdnrslfsInevveifrcpg 278
 QY 223 YEYSERGDAVCTE-SGWRP-LPSCEEKSCDNPYIPN-----GDYSPLRIKHRTG 269
 DB 279 fvmkgprrvkcqalnkwepelpsc-srvccqp--pdvlhaertqrkdnfsp-----g 328
 QY 270 DEITYQCRNGFYPATRGNTAKTSTG-WIP-APRCTLKPCD 308
 DB 329 qevfyscepg-ydlrgaasmrctpggdwspaaptcevkscd 368
 RESULT 14
 AAB53125
 ID AAB53125 standard; Protein; 645 AA.
 XX
 AC AAB53125;
 XX
 DT 28-FEB-2001 (first entry)
 XX

DE Macaca mulatta rhadinovirus 17577 RRV ORF4 protein SEQ ID NO:7.
 XX
 KW genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
 KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
 KW cytosolic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
 KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
 KW splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.
 XX
 OS Macaca mulatta rhadinovirus 17577.
 XX
 PN WO200028040-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26260.
 XX
 PR 06-NOV-1998; 98US-0107507.
 PR 20-NOV-1998; 98US-0109409.
 XX
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 XX
 PI Wong SW, Athelm MK, Searles RP;
 XX
 DR WPI; 2000-376552/32.
 XX
 PT New rhesus rhadino virus for producing non-human primate model useful
 PT for testing potential treatments and efficacy of the candidate vaccine
 PT for conditions associated with RRV infection -
 XX
 PS Claim 5; Page 122-123; 141pp; English.
 XX
 CC The present invention describes a novel rhesus macaque rhadinovirus
 CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
 CC RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
 CC encoded by the genome sequence. The present invention also specifically
 CC claims the individual open reading frame (ORF) nucleotide sequences from
 CC the genome which encode the individual proteins, but these sequences are
 CC not given. A non-human animal infected with RRV can be used for testing
 CC the efficacy of drug in the treatment of condition associated with
 CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative
 CC disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
 CC hypergammaglobulinemia or autoimmune haemolytic anaemia, by
 CC administering the drug to a immuno-compromised non-human primate
 CC preferably Rhesus macaque monkey obtained by as a result of infection
 CC by Simian Immunodeficiency Virus (SIV). RRV is useful for producing
 CC non-human primate model for testing potential treatments for conditions
 CC associated with RRV infection. It is also useful for testing the
 CC efficacy of the candidate vaccine against RRV infection or conditions
 CC associated with its infection by administering the vaccine to the
 CC subject capable of infection with RRV, inoculating the subject with RRV
 CC and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205
 CC to AAB53213 represent sequence used in the exemplification of the
 CC present invention.
 XX
 SQ Sequence 645 AA;

Query Match 18.6%; Score 349.5; DB 21; Length 645;
 Best Local Similarity 25.8%; Pred. No. 1.1e-18;
 Matches 99; Conservative 45; Mismatches 164; Indels 75; Gaps 14;

Qy 1 EDCNELPPRRNTEILTGSWSQDTPEGTQAIYKCRPGYRSLGNVIMV--CRKGWALNP 58
 Db 21 encK--pphftfeyrvksntekdlvsgetaelicrpgyvtntkittetclngtw--stp 76
 Qy 59 LRKCKRCPGHPGTPTFTLTGG-NVFYGVRAVYTCNGEYOLLGEINYNREC---DTD 114
 Db 77 nfpcdkrctpadllngavhnhgdhalkfgsnisyecnegydlgs-nvrfcilqgte 135
 Qy 115 --GWTNDIPICEVYVKCLPVTAPENGKIVSSAMEPDRE-YHFGQAVRFVCSNGYKIEGDEE 171
 Db 136 nvwnsdnspvceiqckcikppavehgdyl-----pnqdvnygydaitfkscisylcivgst 190

Qy 172 MHCSDGFWSEKPKCEVLSCKSPDVING-SPISQKIIYKENERFOYKCNMGVEYSERGD 230
 Db 191 ivctsnkkwsnffptclmivcespqldngyldlglsrryhnghqstvkcsdgnlyvgpet 250
 Qy 231 AVCTESGW-RPLPSC-----EKSQDNFYIPNGDYSPLRI----- 264
 Db 251 lctctnttwvpplpkcvltnnpstmpetpmpetpdygklnlstaktattpnafvtv 310
 Qy 265 -----KURTGDEITYQCRNGFYPATRGNTAKCTSTGWI 297
 Db 311 vspekddvtcvkphferfmvkaendkeysvgasvelicrpgftkmqstvsveclngtw 370
 Qy 298 PAP--RCTLKPCDYPDIKHGGLY 318
 Db 371 tapnakchrkkcptpqellingey 393

RESULT 15
 AAR28557
 ID AAR28557 standard; peptide; 543 AA.
 XX
 AC AAR28557;
 XX
 DT 19-MAR-1993 (first entry)
 XX
 DE CR1-4 (99H, 103E) analogue.
 XX
 KW short consensus repeat; regulator of complement activation;
 KW C3b binding; C4b binding; human complement type 1 receptor.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 1..60
 FT /label= SCR-1
 FT Region 61..122
 FT /label= SCR-2
 FT Region 451..510
 FT /label= SCR-8
 FT Region 511..543
 FT /label= SCR-9
 FT /note= "TRUNCATED"
 FT Misc-difference 99
 FT /note= "Ser substituted by His (SCR-9)"
 FT Misc-difference 103
 FT /note= "Thr substituted by Glu (SCR-9)"
 XX
 PN EP512733-A.
 XX
 PD 11-NOV-1992.
 XX
 PF 28-APR-1992; 92EP-0303826.
 XX
 PR 03-MAY-1991; 91US-0695514.
 XX
 PA (UNIW) UNIV WASHINGTON.
 XX
 PI Atkinson JP, Hourcade D, Krych M;
 XX
 DR WPI; 1992-375009/46.
 XX
 PT Complement activity regulator protein analogues - useful for
 PT treating auto-immune diseases, to suppress transplant rejection,
 PT for diagnosis etc.
 XX
 PS Claim 11; Fig 2 and R11810; 23pp; English.
 XX
 CC The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal
 CC SCR's of CR1. The invention concerns analogues of "regulator of
 CC complement activation" proteins or truncated, hybrid or recombinant
 CC forms of them. CR1-4 is a preferred truncated form and a number of

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 29, 2002, 15:02:17 : Search time 39.15 Seconds
(without alignments)
1050.479 Million cell updates/sec

Title: US-09-316-163-14

Perfect score: 2454

Sequence: 1 EDCKGPPPRENSILSGWS.....DTYYCTENGWSPPKVRIK 428

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2055	83.7	1234	1	NBMSH	complement factor
2	1714	69.8	1231	1	NBHUH	complement factor
3	1709	69.6	449	1	NBHUH	complement factor
4	1242	50.6	669	2	S6551	factor H - bovine
5	813	33.1	452	2	A35068	complement factor
6	811	33.0	303	2	H35068	apolipoprotein H-r
7	569.5	23.2	1053	2	S46199	probable complement
8	566	23.1	808	2	D35069	complement factor
9	529.5	21.6	597	1	S53711	C4BP alpha chain p
10	481.5	19.6	597	1	NBHUC4	C4b-binding protei
11	474.5	19.3	610	1	I46001	C4b-binding protei
12	466.5	19.0	331	2	A45222	complement factor
13	453	18.5	676	2	A45900	complement C3b rec
14	449	18.3	558	2	S79953	C4BP protein alpha
15	447	18.2	2014	2	I36936	complement recepto
16	437	17.8	2489	2	I73012	complement C3b/C4b
17	409	16.7	497	2	JC2054	complement regulat
18	400.5	16.3	482	2	A34924	complement C3b/C4b
19	396	16.1	579	2	A56740	Sperm-egg recognit
20	384.5	15.7	469	1	NBMSG4	C4b-binding protei
21	375	15.3	560	2	I16833	hypothetical prote
22	373	15.2	1025	1	A43526	complement C3d/Eps
23	363	14.8	1091	1	PL0009	complement C3d/Eps
24	336	13.7	661	1	KFHU13	coagulation factor
25	332	13.5	668	2	A46013	coagulation factor
26	319	13.0	343	2	G35070	apolipoprotein H-r
27	318	13.0	440	2	A43519	complement recepto
28	315	12.8	330	2	I56100	complement factor
29	314.5	12.8	263	1	WMVZSP	apolipoprotein H h

ALIGNMENTS

RESULT 1

NBMSH

complement factor H precursor - mouse

N:Alternate names: protein beta-1-H

C:Species: Mus musculus (house mouse)

C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 22-Jun-1999

C:Accession: A26154; I49711; I49728

R:Kristensen, T.; Tack, B.F.

Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986

A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in lengt

A:Reference number: A26154; MUID:86233353

A:Accession: A26154

A:Molecule type: mRNA

A:Residues: 1-1234 <KRI>

A:Cross-references: GB:M2660; NID:g193724; PIDN:AAA37759.1; PID:g387181

R:Natsuume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.

J. Immunol. 144, 358-362, 1990

A:Title: Demonstration of an unusual allelic variation of mouse factor H by the compl

A:Reference number: I49711; MUID:90111033

A:Accession: I49711

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-18 <RES>

A:Cross-references: GB:M31979; NID:g193726; PIDN:AAA37762.1; PID:g193729

R:Munoz-Canoves, P.; Tack, B.F.; Vlk, D.P.

Biochemistry 28, 9891-9897, 1989

A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma

A:Reference number: I49728; MUID:90148935

A:Accession: I49728

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-19 <RE2>

A:Cross-references: GB:J02891; NID:g193805; PIDN:AAA37795.1; PID:g553926

C:Comment: Two codominant alleles of factor H are present in mice.

C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prot

(C5 convertase) in the alternative complement pathway.

C:Genetics:

A:Map position: 1

C:Superfamily: complement factor H; complement factor H repeat homology

C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-1234/Product: complement factor H #status predicted <MPT>

F:21-80/Domain: complement factor H repeat homology <FH01>

F:85-141/Domain: complement factor H repeat homology <FH02>

F:146-205/Domain: complement factor H repeat homology <FH03>

F:210-262/Domain: complement factor H repeat homology <FH04>

F:246-248/Region: cell attachment (R-G-D) motif

F:267-320/Domain: complement factor H repeat homology <FH05>

F:325-385/Domain: complement factor H repeat homology <FH06>

F:389-442/Domain: complement factor H repeat homology <FH07>

F:448-505/Domain: complement factor H repeat homology <FH08>

F:509-564/Domain: complement factor H repeat homology <FH09>

P-selectin precurs
complement factor
P-selectin precurs
membrane cofactor
membrane cofactor
complement control
hypothetical prote
B18L protein - var
P-selectin - rat
membrane cofactor
membrane cofactor
sperm CD46 - human
membrane cofactor
complement control
E-selectin precurs
E-selectin precurs

30 310.5 12.7 768 2 A42755
31 310 12.6 270 2 I37278
32 309.5 12.6 830 2 A30359
33 298.5 12.2 377 2 I54479
34 298.5 12.2 384 2 S01896
35 297.5 12.1 263 1 C36838
36 297.5 12.1 263 2 T28450
37 296.5 12.1 263 2 B72152
38 296 12.1 768 2 I53821
39 294.5 12.0 362 2 JC5194
40 294.5 12.0 369 2 JC5138
41 293.5 12.0 349 2 G02913
42 293.5 12.0 369 2 I57998
43 290.5 11.8 360 2 T42921
44 290 11.8 610 2 A35046
45 272 11.1 612 2 B42755

C:Superfamily: complement factor H; complement factor H repeat homology
 C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-1229/Product: complement factor H #status experimental <MPT>
 F:19-449/Product: complement factor H, short splice form #status experimental <MAT>
 F:21-40/Domain: complement factor H repeat homology <FH01>
 F:85-141/Domain: complement factor H repeat homology <FH02>
 F:146-205/Domain: complement factor H repeat homology <FH03>
 F:210-262/Domain: complement factor H repeat homology <FH04>
 F:246-248/Region: cell attachment (R-G-D) motif
 F:267-320/Domain: complement factor H repeat homology <FH05>
 F:325-385/Domain: complement factor H repeat homology <FH06>
 F:389-442/Domain: complement factor H repeat homology <FH07>
 F:448-505/Domain: complement factor H repeat homology <FH08>
 F:509-564/Domain: complement factor H repeat homology <FH09>
 F:569-623/Domain: complement factor H repeat homology <FH10>
 F:630-684/Domain: complement factor H repeat homology <FH11>
 F:691-744/Domain: complement factor H repeat homology <FH12>
 F:753-803/Domain: complement factor H repeat homology <FH13>
 F:811-864/Domain: complement factor H repeat homology <FH14>
 F:870-926/Domain: complement factor H repeat homology <FH15>
 F:931-984/Domain: complement factor H repeat homology <FH16>
 F:989-1043/Domain: complement factor H repeat homology <FH17>
 F:1048-1102/Domain: complement factor H repeat homology <FH18>
 F:1109-1163/Domain: complement factor H repeat homology <FH19>
 F:1167-1228/Domain: complement factor H repeat homology <FH20>
 F:21-66,52-80,85-129,114-141,146-192,178-205,210-231,237-262,267-309,294-320,325-374,357-379,385-449/Domain: complement factor H repeat homology <FH21>
 F:217/Binding site: carbohydrate (Asn) (covalent) #status absent
 F:529,802,822,882,911/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:718,1029,1095/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 69.88; Score 1714; DB 1; Length 1231;
 Best Local Similarity 67.59; Pred. No. 8,7e-112;
 Matches 289; Conservative 47; Mismatches 92; Indels 0; Gaps 0;

QY 1 EDCKGPPRENSSEILSGSWBOLYSEGTOATYKCRPGYRTLTGTVKYCKGEWVPSNPSR 60
 DB 19 EDCNELPRRNTTEILSGWSBOLYSEGTOATYKCRPGYRTLTGTVKYCKGEWVPSNPSR 78
 QY 61 ICRKRCPGHPGDTFPGFRLAVGSEFFGAKVYVTCDEGYOLLGEIDYRECDADGWTNDI 120
 DB 79 KCQRKCPGHPGDTFPGFRLAVGSEFFGAKVYVTCDEGYOLLGEIDYRECDADGWTNDI 138
 QY 121 PICEVVKCLPTELENGRIVSGAEPDQYFYGQVVRFCNSGPKIEQKEMHCSNGLW 180
 DB 139 PICEVVKCLPTELENGRIVSGAEPDQYFYGQVVRFCNSGPKIEQKEMHCSNGLW 198
 QY 181 SNEKPCQVEISCLPVRVNGDGYLKPVKENERFYKCKGQFYKRGDAVCTGSGWNP 240
 DB 199 SKEAPKVEISCKSPDVINGSPISQKIYKENERFYKCKGQFYKRGDAVCTGSGWNP 258
 QY 241 QPSCEMTCLTPYIPNGIYTPHRIKRIIDDEIRYECNKGFPATRSPVSKCTITGWPAP 300
 DB 259 LPSCEKSCDNPYIPNGIYTPHRIKRIIDDEIRYECNKGFPATRSPVSKCTITGWPAP 318
 QY 301 RCLSLKPCDFQFKGRILYIESRRPYFPVPIGKEYSYSCONGFTTPSGQSYWDLRCVTNG 360
 DB 319 RCLSLKPCDFQFKGRILYIESRRPYFPVPIGKEYSYSCONGFTTPSGQSYWDLRCVTNG 378
 QY 361 WEPEVPLCLRCIFHYVEYGESSYWRORYIEGOSAKVOCHSGYSYLPNGODTYCYTCENGWSP 420
 DB 379 WSPAVCLRCIKCYFPYLENGYNGNQHNGRKFVQOKSIDVACHPGYALPKAQTITVTCMENGWSP 438
 QY 421 PPKCVRIK 428
 DB 439 TPRCIRVK 446

RESULT 3

NBHHS

complement factor H precursor, short splice form [validated] - human

N:Alternate names: complement factor H-related protein; complement protein H
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence_revision 23-Feb-1996 #text_change 08-Dec-2000
 C:Accession: S03013; B60238; A27877; A61103; A26505; S10479
 R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
 Biochem. J. 249, 593-602, 1988
 A:Title: The complete amino acid sequence of human complement factor H.
 A:Reference number: S00254; MUID:88134059
 A:Accession: S03013
 A:Molecule type: mRNA
 A:Residues: 1-449 <RIP>
 A:Cross-references: EMBL:X07523; EMBL:Y00716; NID:g22492; PIDN:CAA30403.1; PID:g75807
 A:Note: part of this sequence, including the amino end of the mature protein was conf
 A:Note: 402-Tyr was also found
 R:Estaller, C.; Schwaebble, W.; Dierich, M.; Weiss, E.H.
 Eur. J. Immunol. 21, 799-802, 1991
 A:Title: Human complement factor H: two factor H proteins are derived from alternativ
 A:Reference number: A60238; MUID:91184292
 A:Accession: B60238
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-33;434-449 <EST>
 A:Note: only portions of this 1.8 kilobase mRNA were sequenced
 R:Schulz, T.F.; Schwaebble, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.
 Eur. J. Immunol. 16, 1351-1355, 1986
 A:Title: Human complement factor H: Isolation of cDNA clones and partial cDNA sequen
 A:Reference number: A27877; MUID:87054207
 A:Accession: A27877
 A:Molecule type: mRNA
 A:Residues: 11-55-401, 'Y', 403-449 <SCH>
 A:Cross-references: GB:X04697; NID:g31991; PIDN:CAB41739.1; PID:g4725976
 A:Note: an additional nucleotide present within the codon for Glu-310 was thought to
 R:Schwaebble, W.; Zwirner, J.; Schulz, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.
 Eur. J. Immunol. 17, 1485-1489, 1987
 A:Title: Human complement factor H: expression of an additional truncated gene produc
 A:Reference number: A61103; MUID:88055295
 A:Accession: A61103
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 27-76 <SC2>
 A:Note: this is a partial sequence of an alternatively spliced 1.8 kilobase mRNA that
 R:Sim, R.B.; Discipio, R.G.
 Biochem. J. 205, 285-293, 1982
 A:Title: Purification and structural studies on the complement-system control protein
 A:Reference number: A26505; MUID:83048213
 A:Accession: A26505
 A:Molecule type: protein
 A:Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>
 R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P
 Biochemistry 31, 3626-3634, 1992
 A:Title: Solution structure of the fifth repeat of factor H: A second example of the,
 A:Reference number: A44551; MUID:92232649
 A:Contents: annotation: NMR structure determination, residues 264-292
 R:Kristensen, T.; Wetzel, R.A.; Tack, B.F.
 J. Immunol. 136, 3407-3411, 1986
 A:Title: Structural analysis of human complement protein H: homology with C4b binding
 A:Reference number: S10479; MUID:86169701
 A:Accession: S10479
 A:Molecule type: mRNA
 A:Residues: 226-401, 'Y', 403-449 <KRI>
 A:Cross-references: GB:M12383; NID:g180472; PIDN:AAA52013.1; PID:g180473
 C:Comment: Factor H has also been found bound to cell membranes in an unknown manner.
 C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liv
 C:Genetics: <HF1>
 A:Gene: GDB:HF1; HF
 A:Cross-references: GDB:120041; OMIM:134370
 A:Map position: 1q32-1q32
 C:Genetics: <HF2>
 A:Gene: GDB:HF2; HF
 A:Cross-references: GDB:129095
 A:Map position: 1q32-1q32
 A:Note: the correspondence between the two loci and the sequences indicated is unclear
 C:Function:

A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also incre
he alternative complement pathway
C:Pathway: complement alternate pathway
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-449/Product: complement factor H, short splice form #status experimental <MAT>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357
F:21/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 69.6%; Score 1709; DB 1; Length 449;
Best Local Similarity 67.4%; Pred. No. 6.5e-112;
Matches 288; Conservative 47; Mismatches 92; Indels 0; Gaps 0;

Qy 1 EDCRPPRESEILSGNSQLYSEGTQATYKCRPGYRTLTGTVYKCKNGEWPSPNSR 60
Db 19 EDCNELPRRTEILTGSNSDQTYPEGTAIYKCRPGYRSLGNVIMVCRKEWALNPLR 78

Qy 61 ICRKPCGHPGDTFGSFLAVGSFEFGAKVYVTCDEGYOLLGIDYRECDADGWTNDI 120
Db 79 KCORPCGHPGDTFGTLTGGNVFYGKAVYTCNGYQLLGEINRECDTGDWTNDI 138

Qy 121 PICEVVKCLPTELEGRIVSGAEPDQYFQGVVRFECNSGFKIEGQKEMHCSEGLW 180
Db 139 PICEVVKCLPTELEGRIVSGAEPDQYFQGVVRFECNSGFKIEGQKEMHCSEGLW 198

Qy 181 SNERPCVIEISCLPRVGDGIYLPVYKENERFQYKCKGQYKRGDAVCTGSGWNP 240
Db 199 SKERPCVIEISCLPRVGDGIYLPVYKENERFQYKCKGQYKRGDAVCTGSGWNP 258

Qy 241 QPSCSEMTCLTPYPNGIYTPHRIKHRIIDDEIRYCKNGFYPATRSPVSKCTITGWIPAP 300
Db 259 LPSCEKSCNDPYPNGDYSPLRIKHRTGDEITTCRNGFYPATRNTAKCTSTGWIPAP 318

Qy 301 RCLSKPCDFPFQKRLYVESRRPFPVPYKGYSCYDNGFTTSPQSYDYLRLCTVNG 360
Db 319 RCTLPCDYPDIKGGGLVHNNRRPFPVPYKGYSCYDNGFTTSPQSYDYLRLCTVNG 378

Qy 361 WEPEVPLRQCIFHYVEGSSYWRRIYEGQSAKVOCHSGYSLPNGODTYCTENGWSP 420
Db 379 WSPAVPLRKLKCYFYLENGYQNHCRKFPVQCKSIDVACHPGYALPKAQTTCENGWSP 438

Qy 421 PPKCVRI 427
Db 439 TPRCIRV 445

RESULT 4
565551
factor H - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S65551
R:Soames, C.J.; Day, A.J.; Sim, R.B.
Biochem. J. 315, 523-531, 1996
A:Title: Prediction from sequence comparisons of residues of factor H involved in the in
A:Reference number: S65551; MUID:96202005
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-669 <SOA>
A:Cross-references: GB:X98697; NID:g1419423; PIDN:CAA67257.1; PID:g1419424
C:Superfamily: complement factor H; complement factor H repeat homology
F:55-114/Domain: complement factor H repeat homology <FH1>

F:296-349/Domain: complement factor H repeat homology <FHR>
F:355-412/Domain: complement factor H repeat homology <FHR>
F:416-471/Domain: complement factor H repeat homology <FHR>
F:476-530/Domain: complement factor H repeat homology <FHR>
F:538-592/Domain: complement factor H repeat homology <FHR>
F:599-651/Domain: complement factor H repeat homology <FHR>

Query Match 50.6%; Score 1242; DB 2; Length 669;
Best Local Similarity 58.4%; Pred. No. 3.8e-79;
Matches 206; Conservative 57; Mismatches 88; Indels 1; Gaps 1;

Qy 76 GSFLAVGSEFEFGAKVYVTCDEGYQLLGEIDYRECDADGWTNDIPICEVVKCLPTELE 135
Db 3 GSPLAEGNQFEGAKVYVTCDEGYQMGEMNFREDTNGWTNDIPICEVVKCLPTEPE 62

Qy 136 NGRIVSGAEPDQYFQGVVRFECNSGFKIEGQKEMHCSEGLWSNKPQCVETSLPP 195
Db 63 NGKIFSDALEPDQYTYGVVQVFECSYMLDGPQIHCISAGVWSAETPKCIVEFCRPP 122

Qy 196 RVENGSGIYLPVYKENERFQYKCKGQYKRGDAVCTGSGWNPQPCSEMTCLTPYIP 255
Db 133 VILNQAVLPKATYKQNERVQYRCAAGFEYQGRGDTVCTKSGWTAPTICETCDPRIP 182

Qy 256 NGIYTPHRIKHRIIDDEIRYCKNGFYPATRSPVSKCTITGWIPAPRCLSKPCDFPFQKHG 315
Db 183 NGVYRPELSKYRGDKITYECKKGFPEIRGTDTATCRDGVVVPKCAWKPCSPYVKHG 242

Qy 316 RLYYESRRPFPVPYKGYSCYDNGFTTSPQSYDYLRLCTVNGWEPEVPLRQCIFHY 375
Db 243 RLYY--SYRGYFPAVYNQGFVYSCDHFVFPVSQRSMDHLACTAEGSPSEPCLRQCIENY 300

Qy 376 VEYGESSYWRRIYEGQSAKVOCHSGYSLPNGODTYCTENGWSPKCVRIK 428
Db 301 LENGHNQREKYLQGEIVRVHCYEGYSLQNDQNTWTCTESGWSPPRCIRVK 353

RESULT 5
A35068
complement factor H-related protein 3A4/5G4 - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 27-Jul-1990 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
C:Accession: A35068; B35068; G35068; D35068; E35068; F35068; G35068
R:Vlk, D.P.; Munoz-Canoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 3193-3201, 1990
A:Title: Identification and sequence analysis of four complement factor H-related tra
A:Reference number: A35070; MUID:90153969
A:Accession: A35068
A:Molecule type: mRNA
A:Residues: 1-452 <Vlk>
A:Cross-references: GB:M29010; NID:g192561; PIDN:AAA37415.1; PID:g387128
A:Note: translation of the nucleotide sequence is not complete
C:Superfamily: complement factor H; complement factor H repeat homology
F:28-81/Domain: complement factor H repeat homology <FH1>
F:86-146/Domain: complement factor H repeat homology <FH2>
F:150-203/Domain: complement factor H repeat homology <FH3>
F:209-266/Domain: complement factor H repeat homology <FH4>
F:270-325/Domain: complement factor H repeat homology <FH5>
F:332-386/Domain: complement factor H repeat homology <FH6>
F:390-451/Domain: complement factor H repeat homology <FH7>

Query Match 33.1%; Score 813; DB 2; Length 452;
Best Local Similarity 71.0%; Pred. No. 2.1e-49;
Matches 137; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

Qy 236 SGWNPQPCSEMTCLTPYIPNGIYTPHRIKHRIIDDEIRYCKNGFYPATRSPVSKCTITG 295
Db 15 TAWLSTARGEKTCSPYILNGIYTPHRIIHRSDDEIRYECNYGYPYVTGSTVSKCTPTG 74

Qy 296 WIPAPRCLSKPCDFPFQKRLYVESRRPFPVPYKGYSCYDNGFTTSPQSYDYLRL 355
Db 75 WIPVPRCTLKPCEFPQFKYGRLYVESLRPNFPVSGNKNYKNSGDSGSPSPSWDYLR 134

A:Cross-references: GB:M29009; NID:q192563; PIDN:AAA37416.1; PID:g309166; GB:J05259

A>Note: translation of the nucleotide sequence is not complete

C:Superfamily: complement factor H; complement factor H repeat homology

F:28-88/Domain: complement factor H repeat homology <FH01>

F:92-145/Domain: complement factor H repeat homology <FH02>

F:151-208/Domain: complement factor H repeat homology <FH03>

F:212-267/Domain: complement factor H repeat homology <FH04>

F:272-325/Domain: complement factor H repeat homology <FH05>

F:332-386/Domain: complement factor H repeat homology <FH06>

F:393-446/Domain: complement factor H repeat homology <FH07>

F:455-505/Domain: complement factor H repeat homology <FH08>

F:511-564/Domain: complement factor H repeat homology <FH09>

F:569-622/Domain: complement factor H repeat homology <FH10>

F:627-681/Domain: complement factor H repeat homology <FH11>

F:688-742/Domain: complement factor H repeat homology <FH12>

F:746-807/Domain: complement factor H repeat homology <FH13>

Query Match 23.1%; Score 566; DB 2; Length 808;

Best Local Similarity 64.5%; Pred. No. 6.8e-32;

Matches 91; Conservative 23; Mismatches 27; Indels 0; Gaps 0;

QY 288 VSKCTITGIPAPRCSLKPCDFPQFKHGRLYEESRRPYFPVPIGKEYSYCDNGFTTFS 347

Db 9 LSNLLTAMLSAKGEVKSCFPQFKYGRLYFEILLRPNFPVPSIGNKYSKCDNGESPPS 68

QY 348 QSYWDYLRCTVNGWEPEVCLRCQIFHYVEYGESSYQWRYYIEGQAKVOCHSGYSLPNG 407

Db 69 GLFWDYLRCTVNGWEPEVCLRCQIFHYVEYGESSYQWRYYIEGQAKVOCHSGYSLPNG 428

QY 408 QDTYYCTGNGWSPPKCVRIK 428

Db 129 QDTWCTGNGWSPPKCVRIK 149

RESULT 9

S53711

C4BP alpha chain precursor - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S53711

Biochim. Biophys. Acta 1261, 285-289, 1995

A:Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence mo

A:Reference number: S53711; MUID:95226458

A:Accession: S53711

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-597 <DEF>

A:Cross-references: EMBL:235490

C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology

F:50-107/Domain: complement factor H repeat homology <FH1>

F:112-169/Domain: complement factor H repeat homology <FH2>

F:174-234/Domain: complement factor H repeat homology <FH3>

F:239-294/Domain: complement factor H repeat homology <FH4>

F:299-360/Domain: complement factor H repeat homology <FH5>

F:364-423/Domain: complement factor H repeat homology <FH6>

F:426-480/Domain: complement factor H repeat homology <FH7>

F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match

Best Local Similarity 21.6%; Score 529.5; DB 1; Length 597;

Matches 148; Conservative 63; Mismatches 187; Indels 117; Gaps 26;

QY 2 DCKGPPPRENSIELSGSWSEQLYSEGTQATYKCRPGYRTLG-TIVKVK-NGEWPVSNPS 59

Db 49 DC-GPPHLLFASSISELSENEYQTILKYTCRPGYRTNLNPLTCKPRLM---SYD 104

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCDTFGSRFLRAGSEFEFGKVVYTCDEGVLLEIDYRECDAD---G 115

Db 105 TFCVKKRCRNPGLPNG--QVEVKTFDFSGQIEFSCSEGYILIGSTT-SHCDIQEKGE 161

QY .60 RICRRPCGHPGCD

Qy	2	CKGPPPPRENSSEILSGWSBQLYSEGTQATYKCRPGV-RTLGTIVKVC-KNGEWPVSNPS	59
Db	49	NCGPPPTLSFAAPMDITLTETRFKGTTLAYTCLPGVVRSHSTQTTLTCSNDGEWYNT--	106
Qy	60	RICRKRCPGHPGDPFGSFRVLAVGSEFEFGAKVYVTCDEGYQLLGEIDYRECDAD----	115
Db	107	-FCYIKKRCRHPGLRNG--QVEIKTDLTSFGSQIEFSCSEGFLLIGSTTSR-CEVQDRGV	162
Qy	116	WTNDIPICEVYVKCLPVTLENGRIVSCAAPDQEYFVGQVVRPECNCSGFKIEQKEMHCS	175
Db	163	WSHPLPOCEIVKCKPPDPIRNGR-----HSGEENFYAGFSVTVYSCDPRFSLGLHASISCT	218
Qy	176	-EN---GLWSNEKPCQVEISCLPPRVENGDIY-LKPVYKENERFOYKCKQGFVYKRGD	230
Db	219	VENETIGWVRSPPTCEKIITCRPDVSHGEVSGFGPIYWKDTIVFKCKQGFVL--RGS	276
Qy	231	AV--C-TGSGWNPO-PSCEWMTCLT-PYPINGIYT--PHRIK---HRIDDEIRYECKNGF	280
Db	277	SVIHCDADSKNWSPPACENSCINLPDIIPHASWETVYPRPTKEDVYVVGTVLRYRCHPGY	336
Qy	281	YPATRSPVSKCTITG--WIPAPRCSLKPCDFPFQFKHGRLYYEESRRP-----	325
Db	337	KPTTDEPTTVICOKNLWTPYQCEALCCPEPKLNGEITQHRKSRPANHCYVYFGDEIS	396
Qy	326	-----YFPVPIG-----KEYSYCYDN	341
Db	397	FSCHETSFRSAICOGDGTSPRTSPCGDCINFPFKTAHGHYKOSSYSFFKEIIECDK	456
Qy	342	GFTTSPGSYWDYLRCYTVNGHEPEVP-CLRQCIFHYVEYGESSYWMORRYIEGOSAKYQCHS	400
Db	457	GYILVGQA---KLSCSYSHWSAPAPOCKALCRKPELVNGLSLVDKQYVEPENVTIQCD	513
Qy	401	GYSLPNCQDYYTYCTEN-GWSPP-PKC	424
Db	514	GYGVV-GPOSITCSGNRTWYPEVPKC	538
RESULT	11		
I46001		C4b-binding protein alpha chain - bovine	
C4b-binding protein alpha chain - bovine		C:Species: Bos primigenius taurus (cattle)	
C:Species: Bos primigenius taurus (cattle)		C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999	
C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999		C:Accession: I46001; S43190	
C:Accession: I46001; S43190		J: Hillarp, A.; Thern, A.; Dahlback, B.	
J: Hillarp, A.; Thern, A.; Dahlback, B.		J: Immunol. 153, 4190-4199, 1994	
J: Immunol. 153, 4190-4199, 1994		A:title: Bovine C4b binding protein. Molecular cloning of the alpha- and beta-	
A:title: Bovine C4b binding protein. Molecular cloning of the alpha- and beta-		A:Reference number: I46001; MUID:95015909	
A:Reference number: I46001		A:Accession: I46001	
A:Accession: I46001		A:Status: preliminary	
A:Status: preliminary		A:Molecule type: mRNA	
A:Molecule type: mRNA		A:Translated from GB/EMBL/DBDJ	
A:Translated from GB/EMBL/DBDJ		A:Residues: 1-610 <HL>	
A:Residues: 1-610 <HL>		A:Cross-references: EMBL:231693; NID:g469117; PID:CA83498.1; PID:g469118	
A:Cross-references: EMBL:231693; NID:g469117; PID:CA83498.1; PID:g469118		C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat	
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat		F:50-107/Domain: complement factor H repeat homology <FH1>	
F:50-107/Domain: complement factor H repeat homology <FH1>		F:112-169/Domain: complement factor H repeat homology <FH2>	
F:112-169/Domain: complement factor H repeat homology <FH2>		F:174-234/Domain: complement factor H repeat homology <FH3>	
F:174-234/Domain: complement factor H repeat homology <FH3>		F:239-294/Domain: complement factor H repeat homology <FH4>	
F:239-294/Domain: complement factor H repeat homology <FH4>		F:299-362/Domain: complement factor H repeat homology <FH5>	
F:299-362/Domain: complement factor H repeat homology <FH5>		F:366-425/Domain: complement factor H repeat homology <FH6>	
F:366-425/Domain: complement factor H repeat homology <FH6>		F:429-483/Domain: complement factor H repeat homology <FHR>	
F:429-483/Domain: complement factor H repeat homology <FHR>		F:487-541/Domain: complement factor H repeat homology <FH8>	
F:487-541/Domain: complement factor H repeat homology <FH8>			
Query Match	19.3%	Score 474.5; DB 1; Length 610;	
Best Local Similarity	27.2%	Pred. No. 1.2e-25;	
Matches 140; Conservative 76; Mismatches 180; Indels 119; Gaps			
Qy	5	GPPIPRENSSEILSGWSBQLYSEGTQATYKCRGXYTLCTIVKCKN-----GEWVPSN	57
Db	51	GIPPYLDFAFPINELNETRFETGTLIRYTCRPGYR-----ISSRKNFLICDGTGDNW---K	102
Qy	58	PSRICRKRCPGHPGDTPFSGFRVLAVGSEPEFGAKVYVTCDEGYQLLGEID-YRECDADG-	115

Db	103	YKEFCVKKRCNCELLNG--QVIVKTDYSFGSEIEFSCSEGVLLIGSANSYCLOQDKGV	160
QY	116	-WTNDIPICEVWCKLPVTELENGRIVSGAAEPOQEYY-FGQVVRFFECNSGFKEIGQKEMH	173
Db	161	VWSDPLPQCIIAKCEPPTISNCRHGG---DEDFYTYGSSVYSCDRDFSMLGKASTS	216
QY	174	CS-EN---GLWSNEKPOCVBISCLPPRVENG---DGILKPVYKENRFQYKCKQGQFVVK	226
Db	217	CRVENKTIIGVWSPPSPCKKVICVQVPWKGKITSG-FGPIVTYQOSIVYACNKGCF--R	272
QY	227	ERGDVAV--C-TGSGWN-POPSCEMTCFL-TPYIPNGIYTPHRTKKHRIIDE-----I	272
Db	273	LEGDSLTHCADNSWNPPTPELNGCLGLPHIPHALW--ERYDHQTQTEQQVYDIDGFVL	330
QY	273	RYECKNGFYATRSPVSKCITIT-----GWLPAKRCSLKPCDRPQF-KHG-RLYYEESRR	324
Db	331	SYKCHFQYKPETDGPT---TJVTQSNLEWSPYTECKVECCPEPLNANYGSITLHRRPSTS	387
QY	325	PYFPVPPIKGYSYYCNG-----F343	
Db	388	THCTYJSGDKISYECHSKYMFDACTKHGTWSPRTPECRPDCSKSPVIAHGQHKVVYSKFF	447
QY	344	TTFSQSQWD-----YLRCTVNGWEPEVP-CLURCIFIHYVEGSSYQWRRYIEG	391
Db	448	TFQHAQVYECDKGYILVGAKELSCTSNGSWSPAVPQCKALCLAKPEIEYGRLSVEKRYVEP	507
QY	392	QSAKVQCHSGSYLPNGODTYTCEN-GWSP-PP-KPC	424
Db	508	EITITQESGYSWV-GSENITCSDRTWYEPKPC	541

```

RESULT      12
A45222
complement factor H-related protein DOWN16 precursor - human
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #accession_revision 30-Sep-1993 #text_change 21-Jul-2000
C:Accession: A45222; S29609
R:Skerka, C.; Kuhn, S.; Gunther, K.; Lingelbach, K.; Zipfel, P.F.
J. Biol. Chem. 268, 2904-2908, 1993
A:Title: A novel short consensus repeat-containing molecule is related to human complement
A:Reference number: A45222; MUID:9315512
A:Accession: A45222
A:Molecule type: mRNA
A:Residues: 1-331 <SKE>
A:Cross-references: EMBL:X58679; NID:g30869; PTID:CAA48639.1; PID:g30870
A:Experimental source: liver
A:Note: sequence extracted from NCBI backbone (NCBIP:124269)
C:Genetics:
A:Gene: GDB:HFL4
A:Cross-references: GDB:l37201
C:Superfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat
C:Keywords: glycoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-331/Product: complement factor H-related protein DOWN16 #status predicted <MAT>
F:23-63/Domain: complement factor H repeat homology <FH01>
F:87-141/Domain: complement factor H repeat homology <FH02>
F:147-204/Domain: complement factor H repeat homology <FH03>
F:211-265/Domain: complement factor H repeat homology <FH04>
F:269-330/Domain: complement factor H repeat homology <FH05>
F:108,186,206,310/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

	Query Match	19.0%;	Score 466.5;	DB 2;	Length 331;
	Best Local Similarity	56.3%;	Pred. No. 2.2e-25;		
	Matches 76; Conservative	18;	Mismatches 40;	Indels	1; Gaps
QY	293	ITGWIPAPRCSLKPCDFPFKGHRLXYEESRRPYEPVPIKEYSYCDNGCTTPSSSYWD	352		
	:	: : : : : :	: : : : : :	: : : : : :	: : : : : :
Db	. 9	LTLWVCANGQVAKCDFDIKHGGLHEHNRPPFYAVGKIYSYCDHFETPSSYWD	68		
QY	353	YLACTVNGNEPEVPLRQCIFHYVEYGESSYMORRYIEGSAKVQCHSGYSLPN-QGDTRY	411		
	:	: : : : : :	: : : : : :	: : : : : :	: : : : : :

```

Db      69 YIHTQNGWSPAVCLRKCFVPILENGYINQYGRKKFVQGNSTEVACHPGYGLPKVRQTTV 128
Oy      412 YCTENGWSPPPKCVR 426
      ||||| 1:1:1
Db     129 TCTENGWSPTPRCIR 143

RESULT 13
A45900
complement C3b receptor type 2 long form precursor - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C:Accession: A45900; I48306
R:Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.
J. Immunol. 144, 3581-3591, 1990
A:Title: The murine complement receptor gene family. IV. Alternative splic
A:Reference number: A45900; MUID:90229754
A:Accession: A45900
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-676 <KUR>
A:Cross-references: GB:M36470
R:Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.
J. Exp. Med. 181, 151-159, 1995
A:Title: Mouse complement regulatory protein Crry/p65 uses the specific me
A:Reference number: I48306; MUID:95105691
A:Accession: I48306
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 21-367 <RES>
A:Cross-references: EMBL:U17128; NID:g595980; PIDN:AAA78271.1; PID:g595982
C:Genetics:
A:Gene: Cr2

```

```

Query Match      18.5%; Score 453; DB 2; Length 676;
Best Local Similarity 28.6%; Pred. No. 4.2e-24;
Matches 126; Conservative 73; Mismatches 17; Gaps 27;

QY 24 YSEGTQATYKCRPGYRTLGTIVKCKNGEWPVNSPRICRKRRCPCGHGDDTFGGSFRLAVG 83
      ::||| ||||| | : : : : : || : : : : : |
Db 42 FAITTWYKCRPGYFRKSFITCLETSKW--SDAQQCFKRKPCMNPOEPLHGSVHINTG 99
      ::||| ||||| | : : : : : || : : : : : |
QY 84 SEFEFGAKVYTCDEGVOLLCEIDYRECDADG---WYNDIPICEVVKCLPYTELENGRIV 140
      ||| : : : : : || : : : : : | : : : : : |
Db 100 --IEFGSTIYSCNOGYRLIGDSSATCIYSDNTVYWNDDNMLPCLCSIPCSPPAISNGDFY 157
      ||| : : : : : || : : : : : | : : : : : |
QY 141 SGAAEPDOEYFGOVVRFCNSG-----FKIEQKEMHCSEN----GLWSENERPQCV-E 189
      | : : : : | : : : : | : : : : | : : : : |
Db 158 SSSRD---SFFYGMVVIYCHTGNREKLPDLVGEKSIYCTSKDNQVGIWNSPPQCIPIR 214
      ||| : : : : | : : : : | : : : : | : : : : |
QY 190 ISCLPPRVENG---DGIYLPVYKENERFOYKCKQGYFYKERGDVAVC--TGSWNQPO--PSC 244
      : : : : | | : : : : | : : : : | : : : : |
Db 215 VKCPMEPIENGLVESG--FKHSFFLNDVTFVKCKSGFTMKGSRIACWQPNKSNPPLPTC 272
      : : : : | | : : : : | : : : : | : : : : |
QY 245 EEMTCLRPY-IPNGIYTPHRIKHRIIDDEIYECKNGEYPATRSPVSKCTIG-WI-PAPR 301
      ||| : : : : | : : : : | : : : : | : : : : |
Db 273 -FWGCLPPQNILHGDYKNDDEFFSGVGOKVSTYTCNPG-YTLIGTLWLVEISLGTWSNTVPT 330
      ||| : : : : | : : : : | : : : : | : : : : |
QY 302 CSLKPCD-FP-QFKHGRLIYEESRRPYPVPIGKEYSVYCDNGFT---TPSOSWYDYLRC 356
      ||| : : : : | : : : : | : : : : | : : : : |

```

Db 331 CEVKSDAIPNHLGRVFLPPN-----LQLGAEVSVFCDLGLQKGRKPS-----QC 378
QY 357 TVNG-----WEPEVPCRLQCFH---YVEYGESSYWRRIEIQSAAKVOCHSGLPNCOD 409
Db 379 IPEGETVIVNNRFPVCEQISCDDPPPEVKNARKPYISLPVPGTVLRYTCSPSYRL-IGEK 437
QY 410 TYVC-TEN---GW-SPPKC 424
Db 438 AIFCISENOVHATDKAPIC 458
RESULT 14
S57953
C4BP protein alpha chain precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 20-Aug-1999
C:Accession: S57953
R: Hillarp, A.; Thern, A.; Dahlback, B.
submitted to the EMBL Data Library, July 1995
A:Description: Molecular cloning of rat C4b-binding protein alpha- and beta-chains: structure and function
A:Reference number: S57953
A:Accession: S57953
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-558 <HIL>
A:Cross-references: EMBL:250051; NID:q899379; PIDN:CAA90391.1; PID:q899380
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F:15-72/Domain: complement factor H repeat homology <FH1>
F:139-199/Domain: complement factor H repeat homology <FH2>
F:204-238/Domain: complement factor H repeat homology <FH3>
F:263-324/Domain: complement factor H repeat homology <FH4>
F:328-386/Domain: complement factor H repeat homology <FH5>
F:390-443/Domain: complement factor H repeat homology <FH6>
F:447-501/Domain: complement factor H repeat homology <FH8>
Query Match 18.3%; Score 449; DB 2; Length 558;
Best Local Similarity 26.7%; Pred. No. 6.5e-24;
Matches 137; Conservative 71; Mismatches 198; Indels 108; Gaps 24;
QY 5 GPPPRENSEILSGSWSEQLSEGTQATYKCRGY-RTLGTIVKVKCN-GEWVSPNSRIC 62
Db 16 GPPDLPALPASENNQDFFSHHTTLRYNCRPGYSRASSOSLYCKPLGKW---QINIA 72
QY 63 RKRPGCHGDPFGSFRSLAVGSEFEFGAKVYVTCDEGYQLLG-EIDYRECDADG--WTND 119
Db 73 VKKSCRNPGLDQNG--KVEVXTDLFLGSQIEFSCSEGYILIGSSTSYCEIOGKGVSWSDP 130
QY 120 IPICEVVKCLPVTELENGRIVSGAAEPDOEYVFGVVRFECSNGPKIEGKEMHCS---- 175
Db 131 LPECVIAKCGMPDITSNKG---HNGREEFEFTYRSVYTKCDPDPFTLLGNASITCTVYVN 187
QY 176 ENGLWSNEKPOCVETSLPPRVENGD-GIYKLPVVKENERGYCKQGVYKERCDAVCT 234
Db 188 TVGWSNPSPPCIERIICWPVKVLTGHTNSGPKHYTKYKDSVRFVCKQGVFLRSGVHICE 247
QY 235 GSG-WNPQPSCEMTCL-TPYIPNG--IYTPHRIK---HRIDDEIRYECKNGFYPATRSP 287
Db 248 ADGWSNPVPCVCELNSCTDIPDIPNAALITSPRKEDYVPGTVLRYICRPGYEPATRQP 307
QY 288 VSKCTI-----TGWIAPRCSLKCPDFQFKHRLYIEESRRP-----YFPVPIGKEYSY 337
Db 308 M---TVICQKDLWSMLRGCKEICCPDPDPKPSVRVQIEKAHPNDCTYF---FGDEVSY 361
QY 338 YCDN-----GFTTPSQSYWDYLR----- 355
Db 362 TCQNDIMLTATKSGDGTWHPTSPCHOSCDPFPPIAHAGRYTKSSY---YVATQVTECEE 419
QY 356 -----CTVNGWEPEVP-CLRQICIFHYVEYGESSYWRRIEIQSAAKVOCHSGYS 403
Db 420 GYRLVGEATISCWTSQWTPAPQCKALCRKPEIGNGVLSNKKQYVETENVTIQDCSGVF 479

QY 404 LPNGODTYVYCTENG-WSP-----PPKCVRI 427
Db 480 M-LGQSITCSENGTWYPKVSRCEQEVPRKCEHV 512
RESULT 15
I36936
complement receptor 1 - chimpanzee (fragment)
C:Species: Pan troglodytes (chimpanzee)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: I36936; I36937
R: Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.
J. Immunol. 153, 691-700, 1994
A:Title: Primary sequence of an alternatively spliced form of CRI. Candidate for the
A:Reference number: I36935; MUID:94292799
A:Accession: I36936
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2014 <BIR1>
A:Cross-references: GB:L24920; NID:g551564; PIDN:AAA51438.1; PID:g557725
A:Accession: I36937
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-397,1751-2014 <BIR2>
A:Cross-references: GB:L24921; NID:g557726; PIDN:AAA51439.1; PID:g557727
C:Genetics: CRI
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: duplication; glycoprotein; tandem repeat
F:18-74/Domain: complement factor H repeat homology <FH01>
F:79-136/Domain: complement factor H repeat homology <FH02>
F:141-207/Domain: complement factor H repeat homology <FH03>
F:213-268/Domain: complement factor H repeat homology <FH04>
F:272-328/Domain: complement factor H repeat homology <FH05>
F:333-391/Domain: complement factor H repeat homology <FH06>
F:396-462/Domain: complement factor H repeat homology <FH07>
F:470-526/Domain: complement factor H repeat homology <FH08>
F:529-586/Domain: complement factor H repeat homology <FH09>
F:723-778/Domain: complement factor H repeat homology <FH10>
F:1041-1107/Domain: complement factor H repeat homology <FH11>
F:1172-1228/Domain: complement factor H repeat homology <FH12>
F:1233-1291/Domain: complement factor H repeat homology <FH13>
F:1296-1362/Domain: complement factor H repeat homology <FH14>
F:1432-1489/Domain: complement factor H repeat homology <FH15>
F:1625-1681/Domain: complement factor H repeat homology <FH16>
F:1749-1815/Domain: complement factor H repeat homology <FH17>
F:1823-1879/Domain: complement factor H repeat homology <FH18>
F:1884-1940/Domain: complement factor H repeat homology <FH19>
Query Match 18.2%; Score 447; DB 2; Length 2014;
Best Local Similarity 30.2%; Pred. No. 3.7e-23;
Matches 137; Conservative 64; Mismatches 165; Indels 88; Gaps 32;
QY 27 GTQATYKCRPGYRTLGTIVKVKCNKEWVSPNSRICRKRCPGDPFGSFRSLAVGSF 86
Db 941 GTSLKYECPYGYRPFISITCLDNLVW--SSPKVCRKSKCTPPDPVNG--MVHVIDI 996
QY 87 EFGAKVYTCDEGYQLLGIEDYRECDADG----WTNDIPICEVVK-LPVTELENGRIVS 141
Db 997 QVGSRIYSCCTGHRLLIGH-SSAECILSGNSAHWSTKPKICQRIPCGLPPT-IANGDFIS 1054
QY 142 GAAEPDOEYVFGVVRFECSNG-----FKTEGKEMHCSN-----GLWSNEKPOC-VEI 190
Db 1055 TNRE---NFHYGVSVYTYRCNLGSRGRKVFELVGEPSIYCTSNDDQGVWSGPAQCII 1111
QY 191 SCLPPRVENG----DGILKVPVYKENERGYCKQGVYKERCDAVCTG--SGWNPQ-PSC 244
Db 1112 KCTPNVNGILVSDN----RSLFSLNVEVFEFCQPGFVYKGVKQALNKEPELFC 1168
QY 245 EEMTCLTPYIPNGIYTP-HRIKHRIDDEIRYECKNGFYPATRSPVSKCTITG-WTP-APR 301

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:12:05 ; Search time 21.93 Seconds
(without alignments)
755.676 Million cell updates/sec

Title: US-09-316-163-14
Perfect score: 2454
Sequence: 1 EDCKGPPPPRENSILSGWS.....DTYYCTENGWSPPPKCVRIK 428

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %	Score	Match	Length	DB ID	Description
1	2055	83.7	1234	1	CFAH_MOUSE	P06909 mus musculus
2	1714	69.8	1231	1	CFAH_HUMAN	P08603 homo sapien
3	1272.5	51.9	685	1	CFAH_BOVIN	Q28085 bos taurus
4	481.5	19.6	597	1	C4BP_HUMAN	P04003 homo sapien
5	478	19.5	330	1	FHR3_HUMAN	Q02985 homo sapien
6	474.5	19.3	610	1	C4BP_BOVIN	Q28065 bos taurus
7	449	18.3	558	1	C4BP_RAT	Q63514 rattus norv
8	437	17.8	2039	1	CR1_HUMAN	P17927 homo sapien
9	384.5	15.7	469	1	C4BP_MOUSE	P06607 mus musculus
10	373	15.2	1025	1	CR2_MOUSE	P19070 mus musculus
11	368	15.0	1033	1	CR2_HUMAN	P20023 homo sapien
12	336	13.7	661	1	F13B_HUMAN	P05160 homo sapien
13	332.5	13.5	331	1	FHR4_HUMAN	Q92496 homo sapien
14	332	13.5	668	1	F13B_MOUSE	Q07968 mus musculus
15	315	12.8	330	1	FHR1_HUMAN	Q03591 homo sapien
16	314.5	12.8	263	1	VCP_YACCV	P10998 vaccinia vi
17	310.5	12.7	768	1	LEM3_MOUSE	Q01102 mus musculus
18	310	12.6	270	1	FHR2_HUMAN	P36980 homo sapien
19	309.5	12.6	830	1	LEM3_HUMAN	P16109 homo sapien
20	298.5	12.2	377	1	MCP_HUMAN	P15529 homo sapien
21	296	12.1	768	1	LEM3_RAT	P98106 rattus norv
22	290	11.8	610	1	LEM2_HUMAN	P16581 homo sapien
23	282.5	11.5	769	1	LEM3_SHEEP	P98109 ovis aries
24	277.5	11.3	611	1	LEM2_CANFA	P33730 canis fami
25	272	11.1	612	1	LEM2_MOUSE	Q00690 mus musculus
26	271.5	11.1	345	1	APOH_MOUSE	Q01339 mus musculus
27	267.5	10.9	345	1	APOH_HUMAN	P02749 homo sapien
28	267.5	10.9	507	1	DAF_CAVPO	Q50401 cavia porce
29	262.5	10.7	345	1	APOH_CANFA	P33703 canis fami
30	256.5	10.5	551	1	LEM2_RABIT	P27113 cryptotagus
31	255	10.4	360	1	CCPH_HSVSA	Q01016 herpesvirus
32	252	10.3	484	1	LEM2_PIG	P98110 sus scrofa
33	246.5	10.0	646	1	LEM3_BOVIN	P42201 bos taurus

RESULT 1

ID	CFAH_MOUSE	STANDARD;	PRT; 1234 AA.
AC	P06909;		
DT	01-JAN-1988 (Rel. 06, Created)		
DT	01-JAN-1988 (Rel. 06, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Complement factor H precursor (Protein beta-1-H).		
GN	HFI OR CFH.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=86233353; PubMed=2940596;		
RA	Kristensen T., Tack B.F.;		
RT	"Murine protein H is comprised of 20 repeating units, 61 amino acids in length.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).		
RN	[2]		
RP	SEQUENCE OF 1-19 FROM N.A.		
RC	STRAIN=BALB/C;		
RX	MEDLINE=90148935; PubMed=2533512;		
RA	Munoz-Canoves P., Tack B.F., Vik D.P.;		
RT	"Analysis of complement factor H mRNA expression: dexamethasone and IFN-gamma increase the level of H in L cells.";		
RL	Biochemistry 28:9891-9897(1989).		
RN	[3]		
RP	SEQUENCE OF 1-18 FROM N.A.		
RX	MEDLINE=90111033; PubMed=2136885;		
RA	Natsume-Sakai S., Nonaka M., Nonaka M., Harada Y.N., Shreffler D.C., Moriaki K.;		
RT	"Demonstration of an unusual allelic variation of mouse factor H by the complete cDNA sequence of the H.2 allotype.";		
RL	J. Immunol. 144:358-362(1990).		
CC	-1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5 CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.		
CC	-1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN MICE.		
CC	-1- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	EMBL; M12660; AAA37759.1; -		
DR	EMBL; J01891; AAA37795.1; -		
DR	EMBL; M31979; AAA37762.1; -		
DR	PIR; A26154; NBMSH.		

34	244.5	10.0	381	1	DAF_HUMAN	P08174 homo sapien
35	243	9.9	390	1	DAF1_MOUSE	Q61475 mus musculus
36	241.5	9.8	958	1	HIG_DROME	Q09101 drosophila
37	238.5	9.7	340	1	DAF_PONPY	P49457 pongo pygma
38	238.5	9.7	345	1	APOH_BOVIN	P17690 bos taurus
39	238	9.7	549	1	LEM2_RAT	P98105 rattus norv
40	236.5	9.6	407	1	DAF2_MOUSE	Q61476 mus musculus
41	221.5	9.0	485	1	LEM2_BOVIN	P98107 bos taurus
42	218	8.9	297	1	APOH_RAT	P26644 rattus norv
43	195	7.9	372	1	LEM1_MOUSE	P18337 mus musculus
44	191.5	7.8	1019	1	LFC_TACTR	P28175 tachypneus
45	190.5	7.8	1019	1	LFC_CARRO	Q26422 carnoscor

ALIGNMENTS

RA MEDLINE-86169701; PubMed-2937845;
RA Kristensen T., Wetsel R.A., Tack B.F.;
RT "Structural analysis of human complement protein H: homology with C4b
RT binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";
RL J. Immunol. 136:3407-3411(1986).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA Bird C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1047-1231 FROM N.A.
RX MEDLINE-91201892; PubMed-1826708;
RA Estaller C., Koistinen V., Schwaible W., Dierich M.P., Weiss E.H.;
RT "Cloning of the 1.4-kb mRNA species of human complement factor H
RT reveals a novel member of the short consensus repeat family related
RT to the carboxy terminal of the classical 150-kDa molecule.";
RN J. Immunol. 146:3190-3196(1991).
RN [6]
RP SEQUENCE OF 19-35.
RX MEDLINE-83048213; PubMed-6215918;
RA Sim R.B., Discipio R.G.;
RT "Purification and structural studies on the complement-system control
RT protein beta 1H (factor H).";
RL Biochem. J. 205:285-293(1982).
RN [7]
RP SEQUENCE OF 1-19 FROM N.A.
RA Vik D.P., Williams S.A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 1-9 FROM N.A.
RA Dominguez O.;
RL Thesis (1993), Hospital Trias I Pujol, Spain.
RN [9]
RP STRUCTURE BY NMR OF 927-985 (SUSHI 16).
RX MEDLINE-91278097; PubMed-1829116;
RA Norman D.G., Barlow P.N., Baron M., Day A.J., Sim B., Campbell I.D.;
RT "Three-dimensional structure of a complement control protein module
RT in solution.";
RN J. Mol. Biol. 219:717-725(1991).
RN [10]
RP STRUCTURE BY NMR OF 264-322 (SUSHI 5).
RX MEDLINE-92232649; PubMed-1533152;
RA Barlow P.N., Norman D.G., Steinkasserer A., Horne T.J., Pearce J.,
RA Driscoll P.C., Sim B., Campbell I.D.;
RT "Solution structure of the fifth repeat of factor H: a second example
RT of the complement control protein module.";
RL Biochemistry 31:3626-3634(1992).
RN [11]
RP STRUCTURE BY NMR OF 866-985 (SUSHI 15 AND 16).
RX MEDLINE-93323119; PubMed-8331663;
RA Barlow P.N., Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P.,
RA Sim B., Campbell I.D.;
RT "Solution structure of a pair of complement modules by nuclear
RT magnetic resonance.";
RL J. Mol. Biol. 232:268-284(1993).
CC -1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
CC C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
CC C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5
CC CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION
CC 341 ONWARD DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; Y00716; CAA68704.1; -;
DR EMBL; X04697; CABA1739.1; ALT_FRAME.
DR EMBL; X07523; CAA30403.1; -;
DR EMBL; M12383; AAA52013.1; -;
DR EMBL; AL049744; CAB70598.1; -;
DR EMBL; M65294; AAA35948.1; -;
DR EMBL; U56979; AAB01987.1; -;
DR EMBL; Z29665; CAA82763.1; -;
DR PIR; S00254; NBHUH.
DR PIR; S03013; S03013.
DR PDB; 1HCC; 15-APR-92.
DR PDB; 1HPH; 15-JUL-93.
DR PDB; 1HFI; 15-JUL-93.
DR MIM; 134370; -;
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 20.
DR SMART: SM00032; CCP; 20.
KW Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW Signal; 3D-structure; Polymorphism; Alternative splicing.
FT CHAIN 19 1231 COMPLEMENT FACTOR H.
FT SIGNAL 1 18
FT DOMAIN 20 81 SUSHI 1.
FT DOMAIN 84 142 SUSHI 2.
FT DOMAIN 145 206 SUSHI 3.
FT DOMAIN 209 263 SUSHI 4.
FT DOMAIN 266 321 SUSHI 5.
FT DOMAIN 324 386 SUSHI 6.
FT DOMAIN 388 443 SUSHI 7.
FT DOMAIN 447 506 SUSHI 8.
FT DOMAIN 508 565 SUSHI 9.
FT DOMAIN 568 624 SUSHI 10.
FT DOMAIN 629 685 SUSHI 11.
FT DOMAIN 690 745 SUSHI 12.
FT DOMAIN 752 804 SUSHI 13.
FT DOMAIN 810 865 SUSHI 14.
FT DOMAIN 869 927 SUSHI 15.
FT DOMAIN 930 985 SUSHI 16.
FT DOMAIN 988 1044 SUSHI 17.
FT DOMAIN 1047 1103 SUSHI 18.
FT DOMAIN 1108 1164 SUSHI 19.
FT DOMAIN 1166 1229 SUSHI 20.
FT DISULFID 21 66 BY SIMILARITY.
FT DISULFID 52 80 BY SIMILARITY.
FT DISULFID 85 129 BY SIMILARITY.
FT DISULFID 114 141 BY SIMILARITY.
FT DISULFID 146 192 BY SIMILARITY.
FT DISULFID 178 205 BY SIMILARITY.
FT DISULFID 210 251 BY SIMILARITY.
FT DISULFID 237 262 BY SIMILARITY.
FT DISULFID 267 309 BY SIMILARITY.
FT DISULFID 294 320 BY SIMILARITY.
FT DISULFID 325 374 BY SIMILARITY.
FT DISULFID 357 385 BY SIMILARITY.
FT DISULFID 389 431 BY SIMILARITY.
FT DISULFID 416 442 BY SIMILARITY.
FT DISULFID 448 494 BY SIMILARITY.
FT DISULFID 477 505 BY SIMILARITY.
FT DISULFID 509 553 BY SIMILARITY.
FT DISULFID 536 564 BY SIMILARITY.
FT DISULFID 569 611 BY SIMILARITY.
FT DISULFID 597 623 BY SIMILARITY.
FT DISULFID 630 673 BY SIMILARITY.
FT DISULFID 659 684 BY SIMILARITY.
FT DISULFID 691 733 BY SIMILARITY.
FT DISULFID 719 744 BY SIMILARITY.
FT DISULFID 753 792 BY SIMILARITY.
FT DISULFID 781 803 BY SIMILARITY.
FT DISULFID 811 853 BY SIMILARITY.
FT DISULFID 839 864 BY SIMILARITY.
FT DISULFID 870 915 BY SIMILARITY.
FT DISULFID 901 926 BY SIMILARITY.
FT DISULFID 931 973 BY SIMILARITY.
FT DISULFID 959 984 BY SIMILARITY.

```
FT DISULFID 989 1032 BY SIMILARITY.
FT DISULFID 1018 1043 BY SIMILARITY.
FT DISULFID 1048 1091 BY SIMILARITY.
FT DISULFID 1077 1102 BY SIMILARITY.
FT DISULFID 1109 1152 BY SIMILARITY.
FT DISULFID 1138 1163 BY SIMILARITY.
FT DISULFID 1167 1218 BY SIMILARITY.
FT DISULFID 1201 1228 BY SIMILARITY.
FT CARBOHYD 529 529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 718 718 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 822 822 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 882 882 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1029 1029 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1095 1095 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 446 449 KTCSS -> SFTL (IN ISOFORM 2).
FT VARSPLIC 450 1231 MISSING (IN ISOFORM 2).
FT VARIANT 402 402 H -> Y.
FT CONFLICT 21 21 /FTID-VAR_001979.
FT CONFLICT 30 30 C -> Q (IN REF. 3).
FT CONFLICT 34 34 T -> V (IN REF. 3).
FT CONFLICT 53 54 T -> Q (IN REF. 3).
FT STRAND 870 871 RP -> IL (IN REF. 2).
FT STRAND 876 876

Query Match 69.8%; Score 1714; DB 1; Length 1231;
Best Local Similarity 67.5%; Pred. No. 1.9e-128;
Matches 289; Conservative 47; Mismatches 92; Indels 0; Gaps 0;

QY 1 EDCCKPPPRENSILGSSSEQLYSGTOATYKCRGYFTLTIVYKVCNGEWFSPNPSR 60
Db 19 EDCNELPPRRTEILLGSWSDQTPGTAQYKCRPGYRSLGNVIMVCRKGWVALNPLR 78
QY 61 ICRKRCGHPGDTFGSFRFLAVGSEPEFGAKVYVTCDEGYQLLGEIDYRECDADGWTNDI 120
Db 79 KCOKRCGHPGDTFGTFTLTGNGVFEYGVKAVYTCNEGYQLLGEINYECDTGDWTNDI 138
QY 121 PICEVVKCLPTELENGRIVSGAAEPDQYFYFGVVFECNSGFKTEGQKMHCSNGLW 180
Db 139 PICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVCNSGYKTEGDEEMHCSDDGF 198
QY 181 SNEKPCVCEISCLPPRVENGDTGLYKPVYKENERFOYKCKQGFVYKERGDAVCTGSGWNP 240
Db 199 SNEKPCVCEISCKSPDVINGSPISOKIYKENERFOYKCMNGEYSESGDAVCTESGWRP 258
QY 241 QPSCEMTCLTPYIPNGIYTPHRIKHRIHDEIRYCKNGFYFATRSPVSKCTITGWIPAP 300
Db 259 LPSCEKSCDNFYIPNGDYSPLRKHRTGDEITYQCRNGFYFATRGNTAKCTSTGNIPAP 318
QY 301 RCLAKPCDFPQKHGLYEESSRRFPYFPICKEYSYCDNGFTTTPSOSYDYLCTVNG 360
Db 319 RCLTKPCDYPDKHGLYHNNRRFPYFPVAVGKYSYCYCDEHFETPSGYSYWDHIHCTQDG 378
QY 361 WEPEVPCRLQCIHFHYVEGESSYWRORYTEGOSAKVQCHSGYSLPNGDYYCTENGWSP 420
Db 379 WSPVPCRLKCYFPYLENGYNQNHGKRFVQGSIDVACHPGYALPRAQTQVTCMENGWSP 438
QY 421 PPKCVRIK 428
Db 439 TPRCIRVK 446

RESULT 3
CFAH_BOVIN
ID CFAH_BOVIN STANDARD; PRT; 685 AA.
AC Q28085;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Complement factor H (H factor 1) (Fragments).
GN HF1.
```

Query Match

51.9%; Score 1272.5; DB 1; Length 685;

```
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 17-685 FROM N.A., AND SEQUENCE OF 1-16.
RC TISSUE=Liver;
RX MEDLINE=96202005; PubMed=8615824;
RA Soames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
in the interaction with complement component C3b.";
RL Biochem. J. 315:523-531(1996).
CC C3b by factor I and also increases the rate of dissociation of the
CC C3bb complex (C3 convertase) and the (C3b)NBB complex (C5
CC convertase) in the alternative complement pathway (By similarity).
CC -1- SIMILARITY: CONTAINS AT LEAST 13 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X98697; CAA67257.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 11.
DR SMART; SM00032; CCP; 11.
KW Complement alternate pathway; Plasma; Repeat; Sushi.
FT NON_TER 1
FT NON_CONS 16
FT DOMAIN <17 67 SUSHI 2.
FT DOMAIN 70 131 SUSHI 3.
FT DOMAIN 134 188 SUSHI 4.
FT DOMAIN 191 246 SUSHI 5.
FT DOMAIN 249 309 SUSHI 6.
FT DOMAIN 311 366 SUSHI 7.
FT DOMAIN 370 429 SUSHI 8.
FT DOMAIN 431 488 SUSHI 9.
FT DOMAIN 491 547 SUSHI 10.
FT DOMAIN 553 609 SUSHI 11.
FT DOMAIN 614 668 SUSHI 12.
FT DOMAIN 675 >685 SUSHI 13.
FT DISULFID 39 66 BY SIMILARITY.
FT DISULFID 71 117 BY SIMILARITY.
FT DISULFID 103 130 BY SIMILARITY.
FT DISULFID 135 176 BY SIMILARITY.
FT DISULFID 162 187 BY SIMILARITY.
FT DISULFID 192 234 BY SIMILARITY.
FT DISULFID 219 245 BY SIMILARITY.
FT DISULFID 250 297 BY SIMILARITY.
FT DISULFID 280 308 BY SIMILARITY.
FT DISULFID 312 354 BY SIMILARITY.
FT DISULFID 339 365 BY SIMILARITY.
FT DISULFID 371 417 BY SIMILARITY.
FT DISULFID 400 428 BY SIMILARITY.
FT DISULFID 432 476 BY SIMILARITY.
FT DISULFID 459 487 BY SIMILARITY.
FT DISULFID 492 534 BY SIMILARITY.
FT DISULFID 520 546 BY SIMILARITY.
FT DISULFID 554 597 BY SIMILARITY.
FT DISULFID 583 608 BY SIMILARITY.
FT DISULFID 615 656 BY SIMILARITY.
FT DISULFID 642 667 BY SIMILARITY.
FT NON_TER 685
SQ SEQUENCE 685 AA; 77536 MW; 69FC9DC8D530E872 CRC64;
```

```

Best Local Similarity   51.28; Pred. No. 1.1e-93;
Matches 219; Conservative 59; Mismatches 91; Indels 59; Gaps 2;

QY 1 EDCGPPPRENSEILSGWSSEQLYSEGTQATYKCRPGYRTLTGTVKVKNGEWPSPSR 60
   ||| |||| :||| |
Db 1 EDCKEPPPRKKEITLSVS----- 18

QY 61 ICRKPCGHPGDTFFGSFRLVAGSEFEGAKVYVTCDEGYQLLGEIDYRECDADGWTNDI 120
   || || :||:||||| ||||| ||||| :||:||||| ||||| ||||| |||||
Db 19 -----GSPHLAEGNQEFYEGAKVYVTCDEGYQVMGEMNFRCDTNGTNDI 63

QY 121 PICEVWKLCPVTELENGRISVGAAPDQEEYFVGQVRFECNSGFKIGQKEMHCSEGLW 180
   ||||| ||||| ||| :| ||||| ||||| ||||| ||||| ||||| ||||| |||
Db 64 PICEVWKLCPVTEPENGKIFSDALEPDQEEYTYGQVQFECNSGYWLDGPKQIHCSAGVW 123

QY 181 SNEKPCQVEISCLPPRVENGDIYLVKPYKENERFYKCKQGFYVYKRGDAVCTGSGWNP 240
   | | :||| | | :| :| :||| ||| :| :| :||| ||| :| :| :||| |||
Db 124 SAETPKCIVEIFCKPPVILNGQAVLPKATYKONERVQYRCAAGFEYQGRDVTCTKSGWTP 183

QY 241 QPSCSEMTCLTPYIPNGIYTPHRIKHRIDDIRVECKNGFYFATRSPVSKCTITGWTAP 300
   ||| ||| ||||| ||| :| :| :||| ||| ||| ||| ||| ||| ||| |||
Db 184 APTCIEITCDPRIPNGVYRPELSYRGQDKITTECKGFFPEIRGTDATCRDGYWVPV 243

QY 301 RCLSKLPKCDFFQFKHGRLYEESRRPYPVPPIGKEYSYVCDNGFTTPSQSYWDYLRCTVNG 360
   ||| :||| ||||| ||| ||| :| :| :||| :| ||| ||| ||| ||| |||
Db 244 RCMKPCSYPIKHGRLYY--SYRGYFAPRVNQGFVYSCDHHFVPPSQRSWDHLACTAEG 301

QY 361 WEPEVPLRCQIFHYHVEYGSSWQRRIYEGQSAKVQCHSGYSLPNQDGTYYCTENQWSP 420
   | | ||||| ||||| :| :| :||| ||| :| :| :||| ||| :| :| :||| |||
Db 302 WSPEEPLRCQIFNYLENGHNHREERYLOGETVRVHCYEGYSLQNDQNTMTCTESQWSP 361

QY 421 PPKCVRIK 428
   ||| :||| :|
Db 362 PPRCIRVK 369

RESULT 4
C4BP_HUMAN STANDARD; PRT; 597 AA.
ID C4BP_HUMAN AC P04003;
DT 23-OCT-1986 (Rel. 02, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein alpha chain precursor (C4bp) (Proline-rich
DE protein) (PRP).
DE C4BPA OR C4BP.
GN GN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP RP
RC TISSUE=Liver;
RX MEDLINE=90073699; PubMed=2590215;
RA Matsuuchi T., Okamura S., Aso T., Sata T., Niho Y.;
RT "Molecular cloning of the cDNA coding for proline-rich protein (PRP):
RT identity of PRP as C4b-binding protein.";
RL Biochem. Biophys. Res. Commun. 165:138-144(1989).
RN [2]
RN SEQUENCE FROM N.A.
RP RP
RX MEDLINE=91113199; PubMed=1989602;
RA Aso T., Okamura S., Matsuuchi T., Sakamoto N., Sata T., Niho Y.;
RT "Genomic organization of the alpha chain of the human C4b-binding
RT protein gene.";
RL Biochem. Biophys. Res. Commun. 174:222-227(1991).
RN [3]
RN SEQUENCE OF 9-81 FROM N.A.
RP RP
RX MEDLINE=88242821; PubMed=3378624;
RA Lintin S.J., Lewin A.R., Reid K.B.M.;
RT "Derivation of the sequence of the signal peptide in human
RT C4b-binding protein and interspecies cross-hybridisation of the C4bp
RT cDNA sequence.";
RL FEBS Lett. 232:328-332(1988).

```

```
DR HSSP; P10998; 1VVD.
DR MIN; 120830; -.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 8.
DR SMART: SM0032; CCP; 8.
DR Complement pathway; Plasma: Glycoprotein; Repeat; Sushi; Signal;
KW Polymorphism.
FT SIGNAL 1 48
FT CHAIN 49 597
FT DOMAIN 49 109
FT DOMAIN 112 171
FT DOMAIN 174 235
FT DOMAIN 238 295
FT DOMAIN 298 361
FT DOMAIN 364 423
FT DOMAIN 425 481
FT DOMAIN 483 539
FT DISULFID 50 96
FT DISULFID 81 108
FT DISULFID 113 154
FT DISULFID 140 170
FT DISULFID 175 217
FT DISULFID 203 234
FT DISULFID 239 281
FT DISULFID 267 294
FT DISULFID 299 348
FT DISULFID 332 360
FT DISULFID 364 387
FT DISULFID 365 409
FT DISULFID 399 422
FT DISULFID 426 468
FT DISULFID 454 480
FT DISULFID 484 525
FT DISULFID 511 538
FT DISULFID 546 558
FT DISULFID 558 588
FT CARBOHYD 221 221
FT CARBOHYD 506 506
FT CARBOHYD 528 528
FT VARIANT 92 92
FT VARIANT 357 357
FT VARIANT 473 473
FT SEQUENCE 597 AA; 67033 MW; 67E03F2EA85A16DD CRC64;

Query Match 19.68; Score 481.5; DB 1; Length 597;
Best Local Similarity 26.58; Pred. No. 6.3e-31;
Matches 134; Conservative 78; Mismatches 195; Indels 99; Gaps 25;

Qy 2 DCKGPPPRENSSEILSGSWSEQLYSEGTOATYKCRPGY-RTLGITIVKVG-KNGEWVPSNPS 59
Db 49 NCGPPPTLSFAAPMDITLTETRFKGTLLKTLCPGYVRSHSTQTLTICNSGGEWYNT-- 106
Qy 60 RICRKRCPGCDPFGSFRVLAVGSEFFGKAVVYTCDEGYQLGEIDYRECDAD----G 115
Db 107 -FCIVKCRHPGELRNG--QVEIKTDLFSGQIEFSCSEGFELIGSTTSR-CEVQDRCVG 162
Qy 116 WTNDIPICEVVKLPVTELENGRIVSGAAEPDQYEQYVQVRFECNSGFKIEGQKEMHCS 175
Db 163 WSHPLQCEIVKCRPPPPDIRNGR----HSGEENFYAYGFVTSYCDPRFSLGHASISCT 218
Qy 176 -EN---GLWSNEKQCQVCEISCLPRVENDGIY-LKPYKENERFOVKCKGQVYKRGD 230
Db 219 VENETIGWRSPPTCEKITCRKPDVSHGEMVSGFGPIYKVDIVFKCQKGFVL--RGS 276
Qy 231 AV--C-TGSGWNPO-PSCEENMTCLT-PYPNGIYT--PHRIK---HRIDDEIRYECKNGF 280
Db 277 SVIHCADSKWNPSPPACENPCINLPDIPHASMETYPRPTKEDVYVVGTVLRYRCHPGY 336
Qy 281 YPATRSPVSKCTITG--WIPAPRCSLAKPDPQPKHGRLYEESRRP----- 325
```

```
Db 337 KPTTDEPTTVICQKNLRWTPYGCALCCPEPKLNNGEITQHRKSRPANHCYVFFYGDEIS 396
Qy 326 -----FPPVPIG-----KEYSYCDN 341
Db 337 FSCHEFSRFAICQDGTWSPRTSPCGDTCNFPPIKIAHGHYKOSYSYFFKEEIIYECDK 456
Qy 342 GFTTFSQSYWMDYLRCVTNGWEPEVP-CLRQCIFHYVEVGESSYWRORYTEGQSAKVOCHS 400
Db 457 GYILVQGA---KLSCSYSHWSAPAPQCKALCKRPVLNGLRSLVDKQDYVEPVNTVIOQDS 513
Qy 401 GYSLPNGQDTYYCTEN-GWSPP-PKC 424
Db 514 GYGVV-GPQSIITCSGNRTWYEPVKC 538

RESULT 5
FHR3_HUMAN
ID FHR3_HUMAN STANDARD; PRT; 330 AA.
AC Q02985; Q9UJ16;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement factor H-related protein 3 precursor (FHR-3) (H factor-like
DE protein 3) (DOWN16).
DE FHR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RC TISSUE=Liver;
RP MEDLINE=93155112; PubMed=8428964;
RA Skerka C., Kuehn S., Guenther K., Lingelbach K., Zipfel P.F.;
RT "A novel short consensus repeat-containing molecule is related to
RT human complement factor H.";
RL J. Biol. Chem. 268:2904-2908(1993).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=20245597; PubMed=10781834;
RA Male D.A., Ormsby R.J., Ranganathan S., Giannakis E., Gordon D.L.;
RT "Complement factor H: sequence analysis of 221 kb of human genomic DNA
RT containing the entire fh, fhr-1 and fhr-3 genes.";
RL Mol. Immunol. 37:41-52(2000).
RN (3)
RP SEQUENCE FROM N.A.
RA Hall R.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBDJ databases.
RN (4)
RP REVIEW.
RX MEDLINE=94226679; PubMed=8172644;
RA Zipfel P.F., Skerka C.;
RT "Complement factor H and related proteins: an expanding family of
RT complement-regulatory proteins?";
RL Immunol. Today 15:121-126(1994).
CC -!- FUNCTION: MIGHT BE INVOLVED IN COMPLEMENT REGULATION.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC -!- SIMILARITY: STRONG, TO FACTOR H.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68679; CAA48639.1; -.
DR EMBL; AL049741; CAB53064.1; -.
DR PIR; A45222; A45222.
DR HSSP; P10998; 1VVD.
```

[illegible]


```

RESULT 8
CRL_HUMAN
ID CRL_HUMAN STANDARD; PRT; 2039 AA.
AC P17927;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement receptor type 1 precursor (C3b/C4b receptor) (CD35
antigen).
GN CRI OR C3BR.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA MEDLINE=89035992; PubMed=2972794;
RA Klickstein L.B., Bartow T.J., Miletic V., Rabson L.D., Smith J.A.,
RA Fearon D.T.;
RT Identification of distinct C3b and C4b recognition sites in the
RT human C3b/C4b receptor (CRI, CD35) by deletion mutagenesis.;
RL J. Exp. Med. 168:1699-1717(1988).
RN [2]
RN SEQUENCE OF 503-2039 FROM N.A.
RP MEDLINE=87168191; PubMed=2951479;
RP Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,
RA Fearon D.T.;
RT "Human C3b/C4b receptor (CRI). Demonstration of long homologous
RT repeating domains that are composed of the short consensus repeats
RT characteristics of C3/C4 binding proteins.";
RL J. Exp. Med. 165:1095-1112(1987).
RN [3]
RN SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.
RP MEDLINE=86067975; PubMed=2933745;
RP Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;
RT "Identification of a partial cDNA clone for the human receptor for
RT complement fragments C3b/C4b.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).
CC -!- FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUKOCYTES, GLOMERULAR
CC PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS, MEDIATES THE
CC BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE
CC ACTIVATED COMPLEMENT.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -!- POLYMORPHISM: CRI IS RESPONSIBLE FOR THE KNOPS BLOOD GROUP SYSTEM.
CC -!- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A
CC LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SRCS OF LHR-A
CC CONTAINED A SITE DETERMINING C4 SPECIFICITY. AND THE N-TERMINAL
CC TWO SRCS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3
CC SPECIFICITY.
CC -!- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CRI.
CC -!- SIMILARITY: CONTAINS 30 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11569; AAA52297.1; -
DR EMBL; M11617; AAA52298.1; -
DR EMBL; M11618; AAA52299.1; -
DR EMBL; Y00816; CAA68755.1; -
DR EMBL; X05309; CAA28933.1; -
DR PIR; A28507; A28507.
DR PIR; A24748; A24748.
DR PIR; B24748; B24748.
DR PIR; C24748; C24748.
DR PIR; S03843; S03843.
DR HSP; P08603; IHP1.
DR MIM; 120620; -

```

```

DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00084; sush1; 30.
DR SMART: SM00032; CCP; 30.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sush1; Blood group antigen.
FT SIGNAL 1 41
FT CHAIN 42 2039 COMPLEMENT RECEPTOR TYPE 1.
FT DOMAIN 42 1971 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1972 1996 POTENTIAL.
FT DOMAIN 1997 2039 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 42 42 PYROLIDONE CARBOXYLIC ACID (POTENTIAL).
FT DOMAIN 42 100 SUSHI A1.
FT DOMAIN 103 162 SUSHI A2.
FT DOMAIN 165 233 SUSHI A3.
FT DOMAIN 237 294 SUSHI A4.
FT DOMAIN 296 354 SUSHI A5.
FT DOMAIN 357 417 SUSHI A6.
FT DOMAIN 420 488 SUSHI A7.
FT DOMAIN 492 550 SUSHI B1.
FT DOMAIN 553 612 SUSHI B2.
FT DOMAIN 615 683 SUSHI B3.
FT DOMAIN 687 744 SUSHI B4.
FT DOMAIN 746 804 SUSHI B5.
FT DOMAIN 807 867 SUSHI B6.
FT DOMAIN 870 938 SUSHI B7.
FT DOMAIN 942 1000 SUSHI C1.
FT DOMAIN 1003 1062 SUSHI C2.
FT DOMAIN 1065 1133 SUSHI C3.
FT DOMAIN 1137 1194 SUSHI C4.
FT DOMAIN 1196 1254 SUSHI C5.
FT DOMAIN 1257 1317 SUSHI C6.
FT DOMAIN 1320 1388 SUSHI C7.
FT DOMAIN 1395 1453 SUSHI D1.
FT DOMAIN 1456 1515 SUSHI D2.
FT DOMAIN 1518 1586 SUSHI D3.
FT DOMAIN 1590 1647 SUSHI D4.
FT DOMAIN 1649 1707 SUSHI D5.
FT DOMAIN 1710 1770 SUSHI D6.
FT DOMAIN 1773 1841 SUSHI D7.
FT DOMAIN 1847 1905 SUSHI E1.
FT DOMAIN 1908 1966 SUSHI E2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.

```



```
FT DISULFID 1066 BY SIMILARITY. 1115
FT DISULFID 1095 BY SIMILARITY. 1132
FT DISULFID 1138 BY SIMILARITY. 1180
FT DISULFID 1166 BY SIMILARITY. 1193
FT DISULFID 1197 BY SIMILARITY. 1240
FT DISULFID 1226 BY SIMILARITY. 1253
FT DISULFID 1258 BY SIMILARITY. 1300
FT DISULFID 1286 BY SIMILARITY. 1316
FT DISULFID 1321 BY SIMILARITY. 1370
FT DISULFID 1350 BY SIMILARITY. 1387
FT DISULFID 1396 BY SIMILARITY. 1439
FT DISULFID 1426 BY SIMILARITY. 1452
FT DISULFID 1457 BY SIMILARITY. 1496
FT DISULFID 1484 BY SIMILARITY. 1514
FT DISULFID 1519 BY SIMILARITY. 1568
FT DISULFID 1548 BY SIMILARITY. 1585
FT DISULFID 1591 BY SIMILARITY. 1633
FT DISULFID 1619 BY SIMILARITY. 1646
FT DISULFID 1650 BY SIMILARITY. 1693
FT DISULFID 1679 BY SIMILARITY. 1706
FT DISULFID 1711 BY SIMILARITY. 1753
FT DISULFID 1739 BY SIMILARITY. 1769
FT DISULFID 1774 BY SIMILARITY. 1823
FT DISULFID 1803 BY SIMILARITY. 1840
FT DISULFID 1848 BY SIMILARITY. 1891
FT DISULFID 1877 BY SIMILARITY. 1904
FT DISULFID 1909 BY SIMILARITY. 1952
FT DISULFID 1938 BY SIMILARITY. 1965
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 959 959 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1028 1028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1310 1310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1481 1481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1540 1540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1605 1605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;

Query Match 17.8%; Score 437; DB 1; Length 2039;
Best Local Similarity 27.7%; Pred. No. 8.5e-27;
Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;

QY 27 GTOATYKCRPGVYRTLTIVKVKNGEMVPSNPSRICRKPCHGPTPGSGRFLAVGSEF 86
DB 66 GYLVNTECRPGVGRPFISICKLSNVTGAKDR--CRRKSCRNPDPVNGVHVIRG--I 121
QY 87 EFGAKVYVTCDEGYQLLIDYRECDADG----WTNDIPICEVVKC-LPVTELENGRIVS 141
DB 122 QFGSQIKYCTKGYRLIGS-SSATCIISGDVTYVNDNETPCDRICPLPT-ITNGDFIS 179
QY 142 GAAEPDQYFGQVVFECNSG-----FKIEGQKEMHCSEN-----GLWSNEKPKQ-VRI 190
DB 180 TNRE----NFHYGSVTVYRCNPGSGGRKRVFELVGEPSIYCTSDNDQVGIWSGPAQCIIN 236
QY 191 SCLPPRVENG----DGIIYLKPYVKENERQYKCKQGFVYKRGDAVCTG-SGWNPO-PSG 244
DB 237 KCTPPNVENGILVSN---RSLSFSLNEVVEFRQCQFGVMKPPRVKRCQALNKWEPLPSC 293
QY 245 EEMTCLTYPVINGIYTPHRIKHRIID----DEIRYECKNGFYPATRSPVSKCIITG-WIP- 298
```

```
Db 294 SRV-CQPP--PDVLHAERTQDKDNFSPGOEVFYSCERG-YDLRGAASMRCTPOGDWSPA 349
QY 299 APRCSLAPCD--FPQFKHGRLYEESRRPYFPV--PIGKEYSYCDNGFTT--PSQSYWD 352
Db 350 APTCEVAKSCDDFMGQLLNGRV-----LFPVNLQLGAKVDFVCEGFKLGSSASY-- 399
QY 353 YLRCTVNG---WEPEVPCLRQCIFHYVEYGESSYMQRRYIEGQSAKVQCHSGYSLPN-- 406
Db 400 ---CVLAGMESLWNSVVPCEQ-----IFCPSPPVIPNGR 431
QY 407 -----GQDTYYCTE-----NG-W-SPPKPC 424
Db 432 HTGKPLEVFPFGKAVNYTCDPHPDRGT*SFDLIGESTIRCTSDPQCGNGVMWSSPAPRC 487

RESULT 9
C4BP_MOUSE
ID C4BP_MOUSE STANDARD; PRT; 469 AA.
AC P08607;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein precursor (C4bp).
GN C4BPA OR C4BP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=88024937; PubMed=3663616;
RA Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;
RT "cDNA structure of murine C4b-binding protein, a regulatory component
of the serum complement system.";
RL Biochemistry 26:4668-4674(1987).
CC -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
(C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
ALSO ACCELERATES THE DEGRADATION OF THE C4B/C2A COMPLEX (C3
CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
-!- SUBUNIT: HOMOPENTAMER; NOT COVALENTLY LINKED. MOUSE LACKS THE
BETA CHAIN OF C4BP.
-!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
-!- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
-!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17122; AAA37312.1; ALT_INIT.
DR PIR; A27117; NBMSC4.
DR HSP; P10998; 1VVD.
DR MGD; MGI:88229; C4bp.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 6.
DR SMART; SM00032; CCP; 6.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT SIGNAL 1 56
FT CHAIN 57 469 C4B-BINDING PROTEIN.
FT DOMAIN 57 116 SUSHI 1.
FT DOMAIN 119 177 SUSHI 2.
FT DOMAIN 180 241 SUSHI 3.
FT DOMAIN 244 300 SUSHI 4.
FT DOMAIN 302 356 SUSHI 5.
FT DOMAIN 358 414 SUSHI 6.
FT DISULFID 58 103 BY SIMILARITY.
FT DISULFID 88 115 BY SIMILARITY.
```

```
FT DISULFID 120 160 BY SIMILARITY.
FT DISULFID 146 176 BY SIMILARITY.
FT DISULFID 181 223 BY SIMILARITY.
FT DISULFID 209 240 BY SIMILARITY.
FT DISULFID 245 287 BY SIMILARITY.
FT DISULFID 273 299 BY SIMILARITY.
FT DISULFID 303 343 BY SIMILARITY.
FT DISULFID 329 355 BY SIMILARITY.
FT DISULFID 359 400 BY SIMILARITY.
FT DISULFID 386 413 BY SIMILARITY.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 428 428 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 469 AA: 51551 MW: 41E137CB8D8C6321 CRC64;

Query Match 15.7%; Score 384.5; DB 1: Length 469;
Best Local Similarity 29.0%; Pred. No. 2.4e-23;
Matches 112; Conservative 64; Mismatches 161; Indels 49; Gaps 21;

QY 5 GPPPRENSILSGWSEQLYSECTQATYKRCPGY-RTLGTIVKVK-NGEWPVSPNSRIC 62
Db 59 GPPPAIPNALPASDVNRTDFESHITLKYELCPYGRGIRNMWVYCKPSEW---EISVSC 115
QY 63 RKRPGCHGDTFPGSLAVGEFEFGAKVVTYCDGYOLLGEIDYRECDAGD---WTN 118
Db 116 AKKHCRNPGYLDNGYV---GETITGSIQIEFSCQEGFILVGS-STSSCEVRKGVAMSN 171
QY 119 DIPICEVVKLPTELENGRIVSGAEPDQYVFGQVRFECNSGRKIEGQKEMHCSENG 178
Db 172 PPPECVIVKCGPPDPSNGK-HSGT---EDFYPNVHGISITCDPGRLVGSFPFIGCTVYN 227
QY 179 ----LMSNEKPCQVEISCLPPRVENGDIY-LKPVYKENERFOYKCKGCFVYKERGDAVC 233
Db 228 KIVPWSSSPPTCEKICISQNLHGVIVSGYKATYTHRDSVRLACLNTGLVLRGVIEC 287
QY 234 TSGG-WNPQSPCEMTCLPYIPNGIYTPHRIKHRIDDEIRYCKNGFYPATRSPYSKCT 292
Db 288 QGNGNSSLPTCEFCDLPPAIVNGYIT-SMYSKI-TLVTVCECDKGYRLVCKAIIIS-CS 344
QY 293 ITGWI-PAPCSLKCDPQFQKHGL-----YEEERRPYFPVPIGKEYSYVCDNGFT-T 345
Db 345 FSKWGTAPCKAL-CQKPEVNGTLDSEKQYVES-----ENVTIQDSGFAML 393
QY 346 PSQSYMDYLRCVNG-WEPEVPCLRQ 370
Db 394 GSQS-----ISCSSEGTWYEVPRCEQ 415

RESULT 10
CR2_MOUSE STANDARD; PRT: 1025 AA.
AC CR2_MOUSE
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Complement receptor type 2 precursor (Cr2) (Complement C3d receptor).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C.
RX MEDLINE-90229735; PubMed-2139457;
RA Fingerth J.D.;
RT "Comparative structure and evolution of murine CR2. The homolog of
the human C3d/EBV receptor (CD21).";
```

```
RL J. Immunol. 144:3458-3467(1990).
RN [2]
RP SEQUENCE OF 12-1025 FROM N.A.
RX MEDLINE-91010789; PubMed-2145366;
RA Molina H., Kinoshita T., Inoue K., Carel J.C., Holers V.M.;
RT "A molecular and immunochemical characterization of mouse CR2.
Evidence for a single gene model of mouse complement receptors 1 and
2.";
RN J. Immunol. 145:2974-2983(1990).
RN [3]
RP SEQUENCE OF 343-401 AND 991-1025 FROM N.A.
RX MEDLINE-89098890; PubMed-2783485;
RA Fingerth J.D., Benedict M.A., Levy D.N., Strominger J.L.;
RT "Identification of murine complement receptor type 2.";
Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).
RN [4]
RP SEQUENCE OF 289-1025 FROM N.A.
RX MEDLINE-89381350; PubMed-2528587;
RA Kurtz C.B., Paul M.S., Aegerter M., Weis J.J., Weis J.H.;
RT "Murine complement receptor gene family. II. Identification and
characterization of the murine homolog (Cr2) to human CR2 and its
molecular linkage to Cr2.";
RN J. Immunol. 143:2058-2067(1989).
CC - FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B
LYMPHOCYTES ACTIVATION.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - TISSUE SPECIFICITY: B LYMPHOCYTES.
CC - SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).
CC - SIMILARITY: CONTAINS 15 SUSHI (SCR) DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; M81083; AAA374451.1; -
EMBL; M35684; AAA37448.1; -
EMBL; M61132; AAA63295.1; -
EMBL; M35685; AAA37450.1; ALT_SEQ.
EMBL; M29281; AAA37447.1; -
PIR; A43526; A43526.
HSSP; P10998; 1VVD.
MGD; MGI:88489; Cr2.
InterPro; IPR000436; Sush1_SCR_CCP.
Pfam; PF00084; sush1; 14.
SMART; SM00032; CCP; 14.
Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
Receptor; Sush1.
SIGNAL 1 11
CHAIN 12 1025 COMPLEMENT RECEPTOR TYPE 2.
DOMAIN 12 963 EXTRACELLULAR (POTENTIAL).
TRANSMEM 964 990 POTENTIAL.
DOMAIN 991 1025 CYTOPLASMIC (POTENTIAL).
DOMAIN 13 74 SUSHI 1.
DOMAIN 81 139 SUSHI 2.
DOMAIN 145 203 SUSHI 3.
DOMAIN 206 264 SUSHI 4.
DOMAIN 267 335 SUSHI 5.
DOMAIN 342 399 SUSHI 6.
DOMAIN 401 458 SUSHI 7.
DOMAIN 462 515 SUSHI 8.
DOMAIN 518 586 SUSHI 9.
DOMAIN 593 649 SUSHI 10.
DOMAIN 653 705 SUSHI 11.
DOMAIN 708 769 SUSHI 12.
DOMAIN 777 834 SUSHI 13.
DOMAIN 840 898 SUSHI 14.
DOMAIN 901 959 SUSHI 15.
DISULFID 14 56 BY SIMILARITY.
DISULFID 42 73 BY SIMILARITY.
```


Db 425 C-NEYLLRGSKISRCBOGKNSPPVC-LEPCTVNVDMNRRNIEMKWKYCKVLHGDLI 482
QY 336 SYCNGGETTPSQSYDWLIRCTVNGWEPEVP-CLRO------CIFHYVEYGESSYW 384
Db 483 DFVCKQGYDLSPLTSLSELSVOCNKEVKYPLCTRKESKGMCTSPPLIKHGVILISST--- 539
QY 385 QRRYIEGOSAKVOCHGSLVPLNGQDTYYCTENGWSPPPKCV 425
Db 540 VDTYENGSSVEYRCFDHDFLEGSREA-YCLDGMWTTPLCL 579

RESULT 13
FHR4_HUMAN
AC Q92496; Q90JY6; STANDARD; PRT; 331 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement factor H-related protein 4 precursor (FHR-4).
GN FHR4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97190290; PubMed=9038172;
RA Skerka C., Hellwege J., Weber W., Tilkorn A., Buck F., Marti T.,
RA Kampen E., Beisiegel U., Zipfel P.F.;
RT "The human factor H-related protein 4 (FHR-4). A novel short consensus
RT repeat-containing protein is associated with human triglyceride-rich
RT lipoproteins.";
RL J. Biol. Chem. 272:5627-5634(1997).
RN [2]
RP SEQUENCE OF 1-19 FROM N.A.
RA Male D.A., Ormsby R.J., Giannakis E., Gordon D.L.;
RT "Promoter region of complement factor H-related 4 (fhr-4) gene.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RX CHARACTERIZATION.
RP MEDLINE=98136603; PubMed=9476126;
RA Hellwege J., Skerka C., Zipfel P.F.;
RT "Biochemical and functional characterization of the factor-H-related
RT protein 4 (FHR-4).";
RL Immunopharmacology 38:149-157(1997).
CC -1- FUNCTION: INVOLVED IN COMPLEMENT REGULATION. CAN ASSOCIATE WITH
CC LIPOPROTEINS AND MAY PLAY A ROLE IN LIPID METABOLISM.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- PTM: GLYCOSYLATED.
CC -1- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: STRONG, TO FACTOR H.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X98337; CAA65980.1; --
CC EMBL; AF190816; AAF05951.1; --
CC HSSP; P10998; 1VVD.
CC MIN; 605337; --
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00084; sushi; 5.
CC SMART; SM00032; CCP; 4.
KW Repeat; Glycoprotein; Sushi; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 331 COMPLEMENT FACTOR H-RELATED PROTEIN 4.
FT DOMAIN 20 84 SUSHI 1.

FT DOMAIN 85 146 SUSHI 2.
FT DOMAIN 147 205 SUSHI 3.
FT DOMAIN 206 266 SUSHI 4.
FT DOMAIN 267 331 SUSHI 5.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 331 AA; 37325 MW; 5A0D04AB4B841424 CRC64;

Query Match 13.5%; Score 332.5; DB 1; Length 331;
Best Local Similarity 44.6%; Pred. No. 2.2e-19;
Matches 62; Conservative 19; Mismatches 53; Indels 5; Gaps 3;

QY 293 ITGWIP-APRCSLKPCDFPQKHRLYYEESRRPVPVPIGKEYSYICNGFTTPSQSYW 351
Db 9 LFLWVSCANGQEVKPCDFPEIQHGLYKSLRLFLPPAAGQSYSYICDNFVTPSGSYW 68
QY 352 DYLRCTVNGWEPEVPCRLQCIFH--YVEYGESSYWMORRYIEGOSAKVCHGSLVPLNGD 409
Db 69 DYIHCTQDGLVTPCLRTCSKSDIEIENGFISSSESVILNKIEIQYKCKPGYATADGNS 128
QY 410 --YYCYCTENGWSPPPKCV 426
Db 129 SGSITCLQNGWSAQPTCIK 147

RESULT 14
F13B_MOUSE
ID F13B_MOUSE STANDARD; PRT; 668 AA.
AC Q07968;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coagulation factor XIII B chain precursor (EC 2.3.2.13) (Protein-
DE glutamine gamma-glutamyltransferase B chain) (Transglutaminase B
DE chain).
OS Mus musculus (Mouse).
GN F13B OR CF13B.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.D2/OSN; TISSUE=Liver;
RX MEDLINE=93224141; PubMed=8468048;
RA Nonaka M., Matsuda Y., Shiroishi T., Moriaki K., Nonaka M.,
RA Natsuume-Sakai S.;
RT "Molecular cloning of the b subunit of mouse coagulation factor XIII
RT and assignment of the gene to chromosome 1: close evolutionary
RT relationship to complement factor H.";
RL Genomics 15:535-542(1993).
CC -1- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,
CC BUT IS THOUGHT TO STABILIZE THE A SUBUNIT AND REGULATE THE RATE
CC OF TRANSGLUTAMINASE FORMATION BY THROMBIN.
CC -1- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER AND KIDNEY.
CC -1- SIMILARITY: CONTAINS 10 SUSHI (SCR) DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D10071; BAA00963.1; --
CC PIR; A46013; A46013.
CC HSSP; P08603; 1HFI.
CC MGI; MGI:88379; F13b.
CC InterPro; IPR000436; Sushi_SCR_CCP.

[illegible]

	Query Match	12.8%	Score 315;	DB 1;	Length 330;
	Best Local Similarity	43.2%;	Pred. No. 5.4e-18;	Mismatches 51;	Conservative 21;
				Gaps 0;	Indels 0;
QY	307	CDFPQFKHGLRYEESRRYPVPICKEYSYCDNGFTTPSQSYWDYLRCITVNGWPEVP	366		
		: : : : : : :			
Db	23	CDFPKINHGLYDEEKPKPSQYPTGEVFVYSCEYNFVSPSKSFWRITTEEGWSPTPK	82		
		: : : : : : :			
QY	367	CLRCIFHYVEYCESSYWQRRTYGOSAKVOCHSGYSLPNGQDTYCTENGWSPPPKC	424		
		: : : : : : :			
Db	83	CLRLCFPFVYENGHSRSGGTHLEGDTVQLICTGYRLQNENNISCVERGWSTPKC	140		
		: : : : : : :			

Search completed: August 29, 2002, 15:12:07
Job time: 583 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:11:41 ; Search time 67.22 Seconds
(without alignments)
1101.485 Million cell updates/sec

Title: US-09-316-163-14

Perfect score: 2454

Sequence: 1 EDCKGPPPRENSEILSGSWS.....DTYYCTENGWSPPPKCVRIK 428

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2426	98.9	1236	11 Q91YB6	Q91YB6 rattus norv
2	1714	69.8	1172	4 Q9NU87	Q9NU87 homo sapien
3	1242	50.6	669	6 Q28085	Q28085 bos taurus
4	813	33.1	452	11 Q61407	Q61407 mus musculus
5	811	33.0	303	11 Q61405	Q61405 mus musculus
6	569.5	23.2	1053	13 Q91275	Q91275 paralabrax
7	566	23.1	808	11 Q61408	Q61408 mus musculus
8	478	19.5	555	11 Q99JAI	Q99JAI cavia porce
9	466	19.0	559	11 Q63135	Q63135 rattus norv
10	453	18.5	679	11 Q99254	Q99254 mus musculus
11	447.5	18.2	522	6 Q28769	Q28769 papio cynoc
12	447	18.2	2014	6 Q29530	Q29530 pan troglod
13	437	17.8	559	4 Q9UGV2	Q9UGV2 homo sapien
14	437	17.8	2039	4 Q16745	Q16745 homo sapien
15	437	17.8	2489	4 Q16744	Q16744 homo sapien
16	436	17.8	1911	6 Q29528	Q29528 papio hamad

ALIGNMENTS

RESULT 1

Q91YB6 ID Q91YB6 PRELIMINARY; PRT: 1236 AA.

AC Q91YB6; DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE COMPLEMENT INHIBITORY FACTOR H.

GN FH.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;

RA Demberg T., Goetze O., Schlaf G.;

RT "Rat complement factor H: molecular cloning, sequencing and expression

in tissues and isolated cells.";

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ320522; CAC67513.1; -

SQ SEQUENCE 1236 AA; 140343 MW; 1AC89FFA28232EBF CRC64;

Query Match 98.9%; Score 2426; DB 11; Length 1236;
Best Local Similarity 99.1%; Pred. No. 3.9e-215;
Matches 424; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 EDCKGPPPRENSEILSGSWSQYSECTQATYKCRPGYRTLTGTIVKVKNGEWPSPSR 60

Db 19 EDCKGPPPRENSEILSGSWSQYSECTQATYKCRPGYRTLTGTIVKVKNGEWPSPSR 78

QY 61 ICRKRCGHPGDPFPFGSFLAVSGSEFEFGAKVYVTCDEGYQLLGEIDYRECDAGWTNDI 120

Db 79 ICRKRCGHPGDPFPFGSFLAVSGSEFEFGAKVYVTCDEGYQLLGEIDYRECDAGWTNDI 138

QY 121 PICEVVKCLPVTLENGRIVSGAEPDQYFYFGVVFECNSGFKIEGQKMHCSENGLW 180

Db 139 PICEVVKCLPVTLENGRIVSGAEPDQYFYFGVVFECNSGFKIEGQKMHCSENGLW 198

Q29531 pan troglod
Q9es77 mus musculu
O08569 cavia porce
Q9wru2 macaca mula
Q28797 pan troglod
Q63612 rattus norv
Q60736 mus musculu
O46545 ovis aries
Q91x48 mus musculu
Q96rm4 homo sapien
Q22328 caenorhabdi
P91658 drosophila
Q92313 mus musculu
Q9dc83 mus musculu
Q9deg0 gallus gall
Q9vyr4 drosophila
O14212 homo sapien
Q28770 papio cynoc
Q13866 homo sapien
Q922h0 mus musculu
Q96qu9 homo sapien
O14006 homo sapien
Q9cut3 mus musculu
Q35520 rattus norv
Q9h284 homo sapien
Q9j2m6 macaca mula
Q91wx0 rattus norv
Q9viu9 drosophila
Q96q03 homo sapien

```
Qy 181 SNEKPOQVEISCLPPRVENGDIYLLKPVYKENERFOYKCKOGFYKRGDAVCTGSGWNP 240
Db 199 SNEKPOQVEISCLPPRVENGDIYLLKPVYKENERFOYKCKOGFYKRGDAVCTGSGWNP 258
Qy 241 OPSCEEMTCLTPYPNGIYTPHRIKRRIDDEIRYCKNGFYPATRSPVSKCTITGWPAP 300
Db 259 OPSCEEMTCLTPYPNGIYTPHRIKRRIDDEIRYCKNGFYPATRSPVSKCTITGWPAP 318
Qy 301 RCSLKPQDFPOFKHRLYYEESRRPYFPVPIGKEYSYCDNGFTTPSOYWDYLRCTVNG 360
Db 319 RCSLKPQDFPOFKHRLYYEESRRPYFPVPIGKEYSYCDNGFTTPSOYWDYLRCTVNG 378
Qy 361 WEPEVPCLRQCFIHYVEYGESSYWRRIYEGOSAKVCHSGYSLPNGODTYCTENGWSP 420
Db 379 WEPEVPCLRQCFIHYVEYGESSYWRRIYEGOSAKVCHSGYSLPNGODTYCTENGWSP 438
Qy 421 PPKCVRIK 428
Db 439 PPKCVRIK 446

RESULT 2
ID Q9NU87 PRELIMINARY; PRT; 1172 AA.
AC Q9NU87;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DJ177p10.1.1 (H FACTOR 1 (COMPLEMENT) ISOFORM 1).
GN HFL
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL049744; CAB70597.1; -.
DR HSSP; P08603; IHFH.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 19.
DR SMART; SM00032; CCP; 19.
SQ SEQUENCE 1172 AA; 132087 MW; 8F5B954C4B4FA454 CRC64;

Query Match 69.8%; Score 1714; DB 4; Length 1172;
Best Local Similarity 67.5%; Pred. No. 2e-149;
Matches 289; Conservative 47; Mismatches 92; Indels 0; Gaps 0;

Qy 1 EDCKGPPPRENSEILSGSWSEQLYSEGTQATYKCRPGYRTLGTTIVKCKNGEWYPSNPSR 60
Db 19 EDCNELPFRNTEILTSWSQDTYPTGTAQYKCRPGYRSIGNVIMVCRKGWALNPLR 78
Qy 61 ICRKRPCGHPDTPFGFRLAVGSEFEFGAKVYTCDEGYQLLGEIDYRECDADGWTNDI 120
Db 79 KCQRPCGHPDTPFGFTLTGGNVFYGKAVYTCDEGYQLLGEINIRYRECDTGDWTNDI 138
Qy 121 PICEVVKCLPVTLENGRIVSGAAEPDQYFGVGVVFECSGFKIEGQKEMHCSENGLW 180
Db 139 PICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVFVCSGFKIEGDEMHCSDDGF 198
Qy 181 SNEKPOQVEISCLPPRVENGDIYLLKPVYKENERFOYKCKOGFYKRGDAVCTGSGWNP 240
Db 199 SKEPKVCISCKSPDVINGSPISQKIYKENERFOYKCNMGVEYSERGDAVCTESGWRP 258
Qy 241 OPSCEEMTCLTPYPNGIYTPHRIKRRIDDEIRYCKNGFYPATRSPVSKCTITGWPAP 300
Db 259 LPCEEKSNDNYPNGDYSPLRKIKHTGDEITYOCRNNGFYPATRGNTAKTSTGTWPAP 318
Qy 301 RCSLKPQDFPOFKHRLYYEESRRPYFPVPIGKEYSYCDNGFTTPSOYWDYLRCTVNG 360
Db 319 RCSLKPQDFPOFKHRLYYEESRRPYFPVPIGKEYSYCDNGFTTPSOYWDYLRCTVNG 378
```

```
Db 319 RCTLKPCDYDPIKHGGLYHENMRRPYFPVAVGKYYSYCYDEHFETPSGSYWDHIHCTQDG 378
Qy 361 WEPEVPCLRQCFIHYVEYGESSYWRRIYEGOSAKVCHSGYSLPNGODTYCTENGWSP 420
Db 379 WSPAVPCLRKCYFFYLENGYNONHGRKRFVQGSIDVACHPGYALPQAQTTVTCMENGWSP 438
Qy 421 PPKCVRIK 428
Db 439 TPCIRVK 446

RESULT 3
ID Q28085 PRELIMINARY; PRT; 669 AA.
AC Q28085;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CCP MODULES 3-12. WITH PARTS OF CCP 2 AND 13 (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=96202005; PubMed=8615824;
RA Seames C.J., Day A.J., Sim R.B.;
RT "Prediction from sequence comparisons of residues of factor H involved
in the interaction with complement component C3b.";
RL Blochem. J. 315:523-531(1996).
DR EMBL; X98697; CAA67257.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 11.
DR SMART; SM00032; CCP; 11.
FT NON_TER 1
FT NON_TER 669
SQ SEQUENCE 669 AA; 75683 MW; D0D9DB30EE747AC2 CRC64;

Query Match 50.6%; Score 1242; DB 6; Length 669;
Best Local Similarity 58.4%; Pred. No. 3.7e-106;
Matches 206; Conservative 57; Mismatches 88; Indels 2; Gaps 1;

Qy 76 GSFRLVGSFEFGAKVYTCDEGYQLLGEIDYRECDADGWTNDIPICEVVKCLPVTLE 135
Db 3 GSPHLAEGNQFEYGAQVYTCDEGYQLLGEIDYRECDADGWTNDIPICEVVKCLPVTLE 62
Qy 136 NGRIVSGAAEPDQYFGVGVVFECSGFKIEGQKEMHCSENGLWSNEKPOQVEISCLPP 195
Db 63 NGKIFSDALEPDQYFGVGVVFECSGFKIEGQKEMHCSAGVWSAETPKCVEIFCKPP 122
Qy 196 RVENGDDIYLLKPVYKENERFOYKCKOGFYKRGDAVCTGSGWNPQSCHEMTCLTPYP 255
Db 123 VILNGQAVLPKATYKQNERVQYRCAAGFEYQGGDTVCTKSGWTPAPTCEITCDDPRIP 182
Qy 256 NGIYTPHRIKRRIDDEIRYCKNGFYPATRSPVSKCTITGWPAPRCSLKPCDFPOFKHG 315
Db 183 NGVYRPELSKRGQDKTYECKKGFPPDIRTDTATCTRDGWWVPVPCRAKWCPSYVIRKH 242
Qy 316 RLYYEESRRPYFPVPIGKEYSYCDNGFTTPSOYWDYLRCTVNGWEPEVPCLRQCFIHY 375
Db 243 RLYY--SYRGYFPAVNVQFVYSCDHHFVPPSQRSWDHLACTAGWSPEEPCLRQCFIHY 300
Qy 376 VEYGESEYWRRIYEGOSAKVCHSGYSLPNGODTYCTENGWSPPPKCVRIK 428
Db 301 LENGHNRREKYLQGETVYRHYEGYSLQNDQNTMTCTESGWSPPPCIRVK 353

RESULT 4
Q61407
```

```
ID Q61407 PRELIMINARY; PRT; 452 AA.
AC Q61407;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE COMPLEMENT FACTOR H-RELATED PROTEIN (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RL Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153969; PubMed=1689298;
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RT "Identification and sequence analysis of 4 complement factor H-related
RT transcripts in mouse liver.";
RL J. Biol. Chem. 265:3193-3201(1990).
DR EMBL; M29010; AAA37415.1; -
DR HSSP; P08603; 1HFI.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 7.
DR SMART; SM00032; CCP; 7.
DR NON_TER 452 452
SQ SEQUENCE 452 AA; 51602 MW; 2B697A4FFC6E13CA CRC64;

Query Match 33.1%; Score 813; DB 11; Length 452;
Best Local Similarity 71.0%; Pred. No. 9.2e-67;
Matches 137; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

QY 236 SGWNPQPSCEMTCLTPYPNGIYTPHRIKHRIIDDEIRYCKNGFYPATRSPVSKCTITG 295
DB 15 TAWLSTAKEEKTCTSPYILNGIYTPHRIIHRSDDEIRYECNRYGYPVGTSTVSKCTPTG 74
QY 296 WIPAPRSLKPCDFPFQKHGRLYEESRRPYPVPPIGKEYSYCDNGFTTPSQSYWDYLR 355
DB 75 WIPVPRCTLKPCFPQFKYGRLYEESLRPNFVSGNKRYSYKCDNGFSPPSGYSWDYLR 134
QY 356 CTVNGWPEVPCLRQCFHYHYVEGSSYWRRYIEGQSAKVCHSGYSLPNGQDTTYCTE 415
DB 135 CTAQGWPEVPCVRKCVFHYVENGDSYWEKIYVQGSCLKVOCYNGYSLQNGQDTTCTE 194
QY 416 NGWSPPPKCVRIK 428
DB 195 NGWSPPPKCIK 207

RESULT 5
ID Q61405 PRELIMINARY; PRT; 303 AA.
AC Q61405;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE COMPLEMENT FACTOR H-RELATED PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153969; PubMed=1689298;
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RT "Identification and sequence analysis of 4 complement factor H-related
RT transcripts in mouse liver.";
RL J. Biol. Chem. 265:3193-3201(1990).
```

```
DR EMBL; M29007; AAA37413.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
SQ SEQUENCE 303 AA; 34498 MW; 51C66E50906F4C24 CRC64;

Query Match 33.0%; Score 811; DB 11; Length 303;
Best Local Similarity 70.8%; Pred. No. 8.5e-67;
Matches 136; Conservative 19; Mismatches 37; Indels 0; Gaps 0;

QY 236 SGWNPQPSCEMTCLTPYPNGIYTPHRIKHRIIDDEIRYCKNGFYPATRSPVSKCTITG 295
DB 15 TAWLSTAKEEKTCTSPYILNGIYTPHRIIHRSDDEIRYECNRYGYPVGTSTVSKCTPTG 74
QY 296 WIPAPRSLKPCDFPFQKHGRLYEESRRPYPVPPIGKEYSYCDNGFTTPSQSYWDYLR 355
DB 75 WIPVPRCTLKPCFPQFKYGRLYEESLRPNFVSGNKRYSYKCDNGFSPPSGYSWDYLR 134
QY 356 CTVNGWPEVPCLRQCFHYHYVEGSSYWRRYIEGQSAKVCHSGYSLPNGQDTTYCTE 415
DB 135 CTAQGWPEVPCVRKCVFHYVENGDSYWEKIYVQGSCLKVOCYNGYSLQNGQDTTCTE 194
QY 416 NGWSPPPKCVRI 427
DB 195 NGWSPPPKCI 206

RESULT 6
ID Q91275 PRELIMINARY; PRT; 1053 AA.
AC Q91275;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE COMPLEMENT REGULATORY PLASMA PROTEIN.
OS Parabrax nebulifer (barred sand bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Seranilidae; Paralabrax.
OX NCBI_TaxID=30873;
RN [1]
RP SEQUENCE FROM N.A.
TX TISSUE=LIVER;
RX MEDLINE=94318039; PubMed=8042982;
RA Dahmen A., Kaidoh T., Zipfel P.F., Gigli I.;
RT "Cloning and characterization of a cDNA representing a putative
RT complement-regulatory plasma protein from barred sand bass (Parabrax
RT nebulifer).";
RL Biochem. J. 301:391-397(1994).
DR EMBL; L21703; AAA92556.1; -.
DR HSSP; P08603; 1HFI.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 16.
DR SMART; SM00032; CCP; 16.
SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D5D3 CRC64;

Query Match 23.2%; Score 569.5; DB 13; Length 1053;
Best Local Similarity 26.7%; Pred. No. 8.2e-44;
Matches 148; Conservative 60; Mismatches 169; Indels 177; Gaps 19;

QY 21 EQLYSEGTOATYKCRPGVRLTGLTVKVCCKNGEWPSPNFSRCRKRKPCGHGPGDTPGSRFL 80
DB 47 EASTPGGRQVRVGNVGS--GFFKLVCVEGKWETRGAK--CQPRSCCHPGDGAQFADFHL 102
QY 81 AVGEFFEGAKVYVTCDEGYQLLGEIDYRECDAGTWNIDPCEVVKCLPVTLENGRIV 140
DB 103 AEGNDFVFGSKVYVTCQGYQMSRINRVCAEGWDCVVPVCSQQC-PLIHVDNNVQV 161
QY 141 SGAAEPDQEEYFGVVRFECSNGPKI-EGQKEMHCSENGLSNEKPKQCEISCLPPRVEN 199
```

```
Db 162 IGPGE---EATGNYVRECKRSEILDSPELYCDERDWSGVPKCKAITCAIPPIEN 218
QY 200 GGIYLVKPVYKENERFOYKCKGFFYKRGDAVCTGSG-----WNPOPSCEEMTCL----- 250
Db 219 GNVPGAIREYKENDVLHYECDRAFKHIDR- PSTCIKQGIKAESWPTPLCESIKCRUTIMD 277
QY 251 -TPYIP-----NGIYT----- 260
Db 278 GTRYEPAYRNLFSPGETLKVICARTSWISTPOETSVVTTQDNGEWSIRPTCOEVRCSNR 337
QY 261 -PHRI-----KHRIIDELRYECKNGFYPATRSPVSKCTITGWIPAP----- 300
Db 338 RPEHVDSDVRSWERYTLDNTRYMCKRGYKRTGGVWTATCGRNGMMPNPLCEVCKTCSKE 397
QY 301 -----RCSLKPCCDFPQFKH 314
Db 398 NIQDAVIVCTQKIYNLNAIKAIYACGEGNRRITLTCGNGWSGDRKCTVKPCPLPKDP 457
QY 315 GRLYEESRRPFPVPIGKEYSYCDNGFTTTPSQSWDYLRCTVNGWEPEVPCLRQCI-- 372
Db 458 NGFF-----RGPY-----TGRVLYYTKDGYKLFTEGHWAEAKVDGVW-----PELTTCISN 505
QY 373 -----PHYVEYGESSYQWRYYIEGSAKVQCHSGYSLPNGQDTYYCTENGW----- 418
Db 506 TTCGKFPPEIPNAEVI---RYPEVQTVQVICNOGYS--TQANSFSCENGWLLYGLSPDQ 560
QY 419 -----SPPPK 423
Db 561 ICTLRADVCGPPPE 574

RESULT 7
ID Q61408 PRELIMINARY; PRT; 808 AA.
AC Q61408:
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE COMPLEMENT FACTOR H-RELATED PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153969; PubMed=1689298;
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RT "Identification and sequence analysis of 4 complement factor H-related
transcripts in mouse liver.";
RL J. Biol. Chem. 265:3193-3201(1990).
DR EMBL; M29009; AAA37416.1; -.
DR HSP; P08603; IHCC.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sush1; 13.
DR SMART; SM00032; CCP; 13.
SQ SEQUENCE 808 AA; 91654 MW; 6FD97D53CE74DF6D CRC64;
```

```
Query Match 23.1%; Score 566; DB 11; Length 808;
Best Local Similarity 64.5%; Pred. No. 1.2e-43;
Matches 91; Conservative 23; Mismatches 27; Indels 0; Gaps 0;
```

```
QY 288 VSKCTITGWIPAPRGLKPCDFPQFKHRLYYEESRRPFPVPIGKEYSYCDNGFTTTPS 347
Db 9 LSNILLTAWLSTAKEVSKCEFPQFKYGLRYEELRPNFPVIGNKYSYKCDNGFSPPS 68
QY 348 OSYWDYLRCTVNGWEPEVPCLRQCIIFHYVEYGESSYQWRYYIEGSAKVQCHSGYSLPNG 407
Db 69 GLFWDYLRCTVGMRPEVPCVRKCVFHYVENGFEAYWEKIIYVQGSGLKVQCYNGISLQNG 128
QY 408 QTYTYCTENGWSPPPKCVRIK 428
```

```
Db 129 QDTMTCTENGWSPPPKCIRIK 149

RESULT 8
ID Q99JA1 PRELIMINARY; PRT; 555 AA.
AC Q99JA1:
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE C4BP ALPHA-CHAIN PRECURSOR.
GN C4BPA.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=EPIDIDYMIS, AND LIVER;
RX MEDLINE=21154058; PubMed=11254714;
RA Nonaka M.I., Wang G., Mori T., Okada H., Nonaka M.;
RT "Novel androgen-dependent promoters direct expression of the C4b-
binding protein alpha-chain gene in epididymis.";
RL J. Immunol. 166:4570-4577(2001).
DR EMBL; AB049465; BAB39737.1; -.
DR EMBL; AB049466; BAB39738.1; -.
DR EMBL; AB049467; BAB39739.1; -.
DR HSP; P10998; LVVD.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sush1; 8.
DR SMART; SM00032; CCP; 8.
KW Signal.
FT SIGNAL 1 13 POTENTIAL.
FT CHAIN 14 555 POTENTIAL.
SQ SEQUENCE 555 AA; 61628 MW; 7AFA5462AFC7B6AC CRC64;
```

```
Query Match 19.5%; Score 478; DB 11; Length 555;
Best Local Similarity 27.2%; Pred. No. 1e-35;
Matches 140; Conservative 73; Mismatches 178; Indels 124; Gaps 27;

QY 5 GPPPRENSIELSGMSEQLYSEGTQATYKCRPGY-RTLGTIVKVK-NGEWMVSNFSRIC 62
Db 16 GPPVLFATPAKELNQETTTNLKTYTCRPGYIRHSDQTLTCKVNGW---RYEVFC 72
QY 63 RRPCHPGDTPFGSFLAVGSEFEFGAKVYTCDSGYQLLG-----EIDYRECDADGW 116
Db 73 SKKQCRNPGDLPHGT--IEVKTDLFLGSKIEFSCSEGNLVGPTTSVCEIHDKGVD--W 127
QY 117 TNDIPICEVVKCLPVTELENGRIVSGAEPDQBYFYGVVVFECNMGFKIEGKEMHCS- 175
Db 128 SVFPFICEIILKCRSPPDISNGK-HSGAVE--DLTYTGSSVYTYSCDPSYLLGNPSISCTV 184
QY 176 ---ENGLWNEKPQCVEISCLPRVE-----NGDGIYLVKPVYKENERFOYKCKGFFYKE 227
Db 185 VAKTVGWSVSPVCKKVICRQPIVQYANLISGFG---PIYFKDTIMFSCQKGFVLKG 240
QY 228 RGDVAV-TGSGWNQPS-CEEMTCL-TPYIPNG-----IYTPHRIKHRIDDEI 272
Db 241 SSLIRCEADNNWNPNPFPVCEPNCSVDIPDIPAYWDSYRRPRKGLYSPGTV-----F 293
QY 273 RYCKNGFYPATRSPVSKCTIT-----GWIPAPRCSLKPCCDFPQFKHGR----- 316
Db 294 KTSCHSGIYPATH---ESTVTVCQDFKWSPPRGCKKVCCEPAPEVNGNSMRADSYSTD 350
QY 317 -----LYEESRRPYF-----PVPIG-----KE----- 334
Db 351 CPFSYNNIFQYRCDSDRQYITSTTCQADGTWKPRVFCGQACHVPPPEIAHGRYREGYLSA 410
QY 335 ---YSYCDNGFTTTPSQSYWDYLRCTVNGWEPEVP-CLROCIFHYVEYGESSYQWRYYIEG 391
Db 411 LSYVVECDGDTLVGQ---NTITCKNSLSSEAPQCAKQCLPKIENGKLSVDKPOYIEP 467
```

```
Qy 392 QSAKVQCHSGYSLPNGQDTYYCTENG-WSP-PPKC 424
   : : | | : : | | : : | | : : | | : : | |
Db 468 ENVTIHCDSGFKL-EGSPSITCEKGTWHGPVKC 501

RESULT 9
ID Q63135 PRELIMINARY: PRT: 559 AA.
AC Q63135;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE COMPLEMENT REGULATORY PROTEIN.
GN CRRY.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=96006570; PubMed=7590969;
RA Ouligg R.J., Lo C.F., Alexander J.J., Sneed A.E., Moxley G. III;
RT "Molecular characterization of rat Crry: widespread distribution of
RT two alternative forms of Crry mRNA.";
RL Immunogenetics 42:362-367(1995).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Dohi N., Sakurada C., Nonaka M., Okada N., Okada H.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; L36532; AAA91821.1; -.
DR EMBL; D42115; BAA22548.1; -.
DR HSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 7.
DR SMART; SM00032; CCP; 7.
DR SEQUENCE 559 AA; 61680 MW; 29E10F6A21DB9B6E CRC64;

Query Match 19.0%; Score 466; DB 11; Length 559;
Best Local Similarity 26.6%; Pred. No. 1.3e-34;
Matches 122; Conservative 91; Mismatches 190; Indels 56; Gaps 26;

Qy 3 CKGPPPRENSILSGSWSEQLYSGTQATYKCRGY--RTLGTIVKVKNGEWPSPNSR 60
   | | : : : : | | : : | | : : | | : : | |
Db 38 CPAPPLFPYAKPIPT-DESTFPVGTSLKYECRPGYIKRQFSITCEV--NSVW--TSPOD 92

Qy 61 ICRKRCGHPGDTFGSFLAAGSEFEFGAKVVTCDGTYLLGEIDYRECDAD---GWT 117
   | : : : | | : : : : | | : : | | : : | | : : | |
Db 93 VCIRKQETPLDPQNGI--VHVNTDIRGSSITTCNBYRLGSSSAMCIISQSVAMD 150

Qy 118 NDIPICEVVKCLPYTELENGRIVSGAEPDQE-YFGQVVRFECSG-----FKIEGQK 170
   : : : : | | : : | | : : | | : : | | : : | |
Db 151 AEAPICESIPCEIPPSIPNGDFFS-----PNREDFHYGMVVTYQCNTDARGKLENLVCEP 206

Qy 171 EMHCS-----ENGLSNKEPKQCVELS-CLPPRVNGDGIYL-KPVYKENERQYCKQGFV 224
   : : | | : : | | : : | | : : | | : : | | : : | |
Db 207 SIHCTSIDGGQGVGWSGPPPPQCIELNCKTPPHVENAVIVSKNLSFLSLDMVFEFCQDGF 266

Qy 225 YKRGDAVCTG-SQWNPQ-PSCEMTCLTPV---IPNG-IYTPHRIKHRIIDEIRYECKN 278
   | | : : | | : : | | : : | | : : | | : : | |
Db 267 MKGSSVYCRSLNREWPQLPCFVKVSGAGFLGPNHGVFPQNL--OLGAKVTFVCNT 324

Qy 279 GFYPATSPVSKCTITG----W-IPAPRCSLKPCDFQFKHGRLYEESRRPYFPVPIGK 333
   | | : : | | : : | | : : | | : : | | : : | |
Db 325 G-YLKGNSSSHCVLDGVESINWSSVPCEQVICKLPQDMSGFGKGLQMKDYY---YGD 380

Qy 334 EYSYCDNGFTTSPQSWDYLRCTVN-GWEPEVP-CLRQCIFHYVE-----YGESSYQWR 387
   : : : : | | : : | | : : | | : : | | : : | |
Db 381 NVALECEDGTYTLEGG---OSQCOSDASWDPLPKCVSQVICKLPQDMSGFQKGLQMKD 437
```

```
Qy 388 YIEQSAKVQCHSGYSLPNGQDTYYCTENGWSP-PPKC 425
   | | : : | | : : | | : : | | : : | |
Db 438 YYYGDNVALECEDGYTLEGSQSQCQSDASMDPLPKCV 476

RESULT 10
ID Q99254 PRELIMINARY: PRT: 679 AA.
AC Q99254;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=90229754; PubMed=2139460;
RA Kurtz C.B., O'Toole E., Christensen S.M., Weis J.H.;
RT "The murine complement receptor gene family. IV. Alternative splicing
RT of Cr2 gene transcripts predicts two distinct gene products that share
RT homologous domains with both human CR2 and CR1.";
RL J. Immunol. 144:3581-3591(1990).
[2]
RN [2]
RP SEQUENCE OF 21-367 FROM N.A.
RX MEDLINE=95105691; PubMed=7528766;
RA Kim Y.O., Kinoshita T., Molina H., Hourcade D., Seya T., Wagner L.M.,
RA Holers V.M.;
RT "Mouse complement regulatory protein Crry/p65 uses the specific
RT mechanisms of both human decay-accelerating factor and membrane
RT cofactor protein.";
RL J. Exp. Med. 181:151-159(1995).
DR EMBL; U17128; AAA78271.1; -.
DR EMBL; U17123; AAA78271.1; JOINED.
DR EMBL; U17124; AAA78271.1; JOINED.
DR EMBL; U17125; AAA78271.1; JOINED.
DR EMBL; U17126; AAA78271.1; JOINED.
DR EMBL; U17127; AAA78271.1; JOINED.
DR EMBL; M36470; AAA37449.1; -.
DR HSP; P10998; 1VVD.
DR MGD; MGI:88489; Cr2.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 10.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 10.
DR NON_TER 679 679
SQ SEQUENCE 679 AA; 74916 MW; 52FC00FDCED20CDC CRC64;

Query Match 18.5%; Score 453; DB 11; Length 679;
Best Local Similarity 28.6%; Pred. No. 2.7e-33;
Matches 126; Conservative 73; Mismatches 178; Indels 64; Gaps 27;

Qy 24 YSEGTQATYKCRPGVYRTLTGTVKVKNGEWPSPNSRICRKRCPGDPGTFGSFLAVG 83
   : : | | : : | | : : | | : : | | : : | |
Db 42 FAIGTTHYKCRPGVFRKSFITCLETSKW--SDAQQCFKRPCKMNPQPLHGSVHINTG 99

Qy 84 SEFEFGAKVVTCDGTYLLGEIDYRECDADQ---WTNDIPICEVVKCLPYTELENGRIV 140
   : : | | : : | | : : | | : : | | : : | |
Db 100 --IEFGSTITYSCNGYRLIGDSSATCIVSDNTVMNDMPICESIPCSPPAISNGDFY 157

Qy 141 SGAAEPDQDYFQGVVRFECSG-----FKIEGQKEMHCSEN-----GLWSNEKQCV-E 189
   | : : : | | : : | | : : | | : : | | : : | |
Db 158 SSSRD---SFFYGMVVTYTYCHTKNRKRLFDLVGKSIYCTSKDNQGVGWNPPQCIPR 214

Qy 190 ISCLPPRVENG---DGIYKPKVYKENERQYKCKQGFYKRGDAVC-TGSGWNI-Q-PSC 244
   | : : : | | : : | | : : | | : : | | : : | |
Db 215 VKCPMEIENGLVESG---FKHSFFLNDTVIFKCKSGFTMKGSRIAWCPNKSPPPLPTC 272
```

```
QY 245 EEMTCLTPY-IPNGIYTPHRIKRIHDEIRYCKNGFYPATRSPVSKCTITG-WI-PAPR 301
DB 273 -FMGLCPONILHGNDYKNDKDEFFSVGOKVSYTCNPG-YTLTGTLNVECTSLGTSWNTVPT 330
QY 302 CSLKPCD-FP-OFKNGRLYYESRRPYFPVPIGKEYSYVCDNGFT---TPSQSYWDYLR 356
DB 331 CEVKSCDAIPNHLHGRVFLPN-----LQIGAESVFVCDLGFQKLGKPS-----QC 378
QY 357 TVNG---WEPEVPCLRQCIFH---YVEVGESSYWMORRYIEGQSAKVQCHSGYSLPNGOD 409
DB 379 IPEGETVIMNNKFPVCEQISCDPPPEVNNARKPYISLPVPGTVLRYKTCSPSYRL-IGEK 437
QY 410 TYIC-TEN---GW-SPPPKC 424
DB 438 AIFCISENQVHATWDKAPIC 458

RESULT 11
Q28769 ID Q28769 PRELIMINARY; PRT; 522 AA.
AC Q28769;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).
OS Papio cynocephalus (Yellow baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Papio.
OX NCBI_TaxID=9556;
RN [1]
RP SEQUENCE FROM N.A.
RA Birmingham D.J., Lodar C.M., Shen X.-P., Chen W.;
RC TISSUE=BONE MARROW;
RT "The baboon erythrocyte complement receptor is a glycoprophosphatidyl
RT inositol-linked protein encoded by a homologue of the human CR1-like
RT genetic element."
RL SUBMITTED (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L77977; AAA99004.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 7.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 7.
KW Receptor.
FT NON_TER 1
SQ SEQUENCE 522 AA; 56626 MW; 312FCBE03ADF19DC CRC64;
```

```
Query Match 18.2%; Score 447.5; DB 6; Length 522;
Best Local Similarity 29.7%; Pred. No. 6.2e-33;
Matches 134; Conservative 66; Mismatches 168; Indels 83; Gaps 31;

QY 27 GTQATYKCRPGVYTLGTIVKCKNGEWPNPSRICKRKPCHGPDTPFGSFLAVGSEF 86
DB 71 GTLYIECLPGYHGKPFIIKLSNWSVTSKDK---CTRKSCRNPDPVNG--MVHVIKDI 126
QY 87 EFGAKVVTCDGEGYOLLGEIDYRECDADG---WTNDIPICEVVKC-LPVTLENGRIYS 141
DB 127 QFGSQINSCNKGRLIGS-SSATCIISGNVIVNDNETPICEIIPCGLPPT-IANGDFIS 184
QY 142 GAAEPDQYVY-FGQVVFECNSG-----FKIEGQKEMHCSEN-----GLWSNEKPCQ-VE 189
DB 185 ----TSREYFPGSVVTVYRCNLGSRKKLFLVLGEPISYICTSKDDQVGIWSGAPQCIIIP 240
QY 190 ISCLPPRVENGDIYL-KPVYKENERFOYKCKGQGVYKRGDAVCTG-SGNWPO-PSCEE 246
DB 241 NKCMPPNVENGVLVSVNSRSLFSLNEWEFCQPGFVMKGRPHRVQCALNKWEPELPCSR 300
QY 247 MTCLTPYIPNGIYTPHRIKRIHDEIRYCKNGFYPATRSPVSKCTITG-WIP-APRCSL 304
```

```
DB 301 VCQPPEILLGHEHTPSHQDFSPQGEVYFSCPEG-YDLRGAASLHCTPQGDWNPAPICTV 359
QY 305 KPCD--FPQFKNGRLYYESRRPYFP--VPICKEYSYVCDNGFTTFSQSYWDYLRCTVNG 360
DB 360 KSCDDFLGQLPHGRV-----LFPNLQLGAKVSVFVCDGEPRLKGRF---ASHCVLAG 408
QY 361 ----WEPEVPCLRQCIF-----HV-VEYGESSYWMORRYIEGQSAKVQCHSGY 402
DB 409 MKALNNSVPVCEQ-IFCNPAPAILNGRHIGAPLDIPY-----GKEYSYICDHPDR 460
QY 403 SLP---NGQDTYYCTE-----NG-W-SPPPKC 424
DB 461 GMTVNLIGESTIRCTSDPOGNGVWSSPAPRC 491

RESULT 12
Q29530 ID Q29530 PRELIMINARY; PRT; 2014 AA.
AC Q29530;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).
GN CR1.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94292799; PubMed=8021505;
RA Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;
RT "Primary sequence of an alternatively spliced form of CR1. Candidate
RT for the 75,000 Mr) complement receptor expressed on chimpanzee
RT erythrocytes."
RL J. Immunol. 153:691-700(1994).
DR EMBL; L24920; AAA51438.1; -.
DR HSSP; P08603; 1HFI.
DR InterPro; IPR001424; SOD_CU_ZN.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR000834; Zn_carboOpept.
DR Pfam; PF00084; sushi; 30.
DR SMART; SM00032; CCP; 30.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW Receptor.
FT NON_TER 1
SQ SEQUENCE 2014 AA; 221281 MW; 6D6C3A74D81F1DB9 CRC64;
```

```
Query Match 18.2%; Score 447; DB 6; Length 2014;
Best Local Similarity 30.2%; Pred. No. 3.8e-32;
Matches 137; Conservative 64; Mismatches 165; Indels 88; Gaps 32;

QY 27 GTQATYKCRPGVYTLGTIVKCKNGEWPNPSRICKRKPCHGPDTPFGSFLAVGSEF 86
DB 941 GTSLAYECRPEYGRPFSTICLDNLVW--SSPKDVCRRKSKCTPPDPVNG--MVHVIKDI 996
QY 87 EFGAKVVTCDGEGYOLLGEIDYRECDADG---WTNDIPICEVVKC-LPVTLENGRIYS 141
DB 997 QVGSRIYNSCTTGHRLIGH-SSAECILSGNSAHWSTKPKICQRIQICGLPPT-IANGDFIS 1054
QY 142 GAAEPDQYVYGOVVFECNSG-----FKIEGQKEMHCSEN-----GLWSNEKPCQ-VEI 190
DB 1055 TNRE---NFHYGSVVTVYRCNLGSRKKRKFVYELVGEPSIYICTSNDQVGIWSGAPQCIIIPN 1111
QY 191 ISCLPPRVENGDIYL-KPVYKENERFOYKCKGQGVYKRGDAVCTG-SGNWPO-PSC 244
DB 1112 KCTPNVENGILVSDN---RSLFSLNEWEFCQPGFVMKGRPHRVQCALNKWEPELPC 1168
QY 245 EEMTCLTPYIPNGIYTPHRIKRIHDEIRYCKNGFYPATRSPVSKCTITG-WIP-APR 301
```


DR	InterPro; IPR001424; SOD_CU_ZN.
DR	InterPro; IPR000436; Sushi_SCR_CCP.
DR	InterPro; IPR000834; Zn_carbopept.
DR	Pfam; PF00084; sushi; 30.
DR	SMART; SM00032; CCP; 30.
DR	PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR	PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW	Receptor.
SQ	SEQUENCE 2039 AA; 223603 MW; B82FCB1LC6BI6635 CRC64;

Query Match	17.8%; Score 437; DB 4; Length 2039;
Best Local Similarity	27.7%; Pred. No. 3.2e-31;
Matches 132; Conservative	63; Mismatches 149; Indels 132; Gaps 31;

Qy	27 GTQATYKCRPGYRTLGIIVKKNGEWPSPNSRICRKRPCGHGDDTDFGFRILAVGEF 86
Dd	66 GTYLNEICRPYSGRPFSSIICLKSVMTGAKDR--CRKSKSRNPDPVNGVMHVTKG--I 121
Qy	87 EFGAKVVYTCDEGVQLLEIDYRECDADG----WTNDIPICEVKC-LPVTLENGRIVS 141
Dd	122 QFGSQIKYSCTKGRLIGS--SSAICIISGDVIWNETPICDRIPCGLPPT-ITNGDPI 179
Qy	142 GAAPEDQEYYVQGVVRFECSNG-----FKIEGOKEMHCSEN-----GLWSNEKPQC-VEI 190
Dd	180 TNRE---NFHYGSVVTVRCNPGSGRKVFELVGPSIVCTSDNQDVQVLSGPAQCIIPN 236
Qy	191 SCLPFRVNG---DGIIYLKPYVENRFQYKCKQGFFYKERGDAVCTG-SGNWPK-PSC 244
Dd	237 KCTPNVENGILVSDN---RSFLSLNVVEFERCGFVFMKGRPVKCOALNKWEPLPSC 293
Qy	245 EEMTCLTPYIPNGYITPHRIKHRIID---DBIRYECKNGFYPATRSPVSKCTITG-WIP- 298
Dd	294 SRV-COPP--PDVLHAERTORDKNFSPGQEVFYSCPG-YDLRGAAAMRCTPOGDHSPA 349
Qy	299 APRCSLPACD--FPQFKHGRIYYEBSRRPPYPV--PTGKEYSYVDNGFTT--PSQSYWD 352
Dd	350 APTCEVSCDDFMGOLLNGRV-----LPPVNLQGAKYDYFCDEGFQLKGSASV-- 399
Qy	353 YLRCTVAG---WBEPVPCLRQCIHFHYVEGESSYWORRYTEGQSAKVQCHSGYSLNP-- 406
Dd	400 ---CVLAGMESLNSSVPVCQ-----;-----IFCPSPVIPNGR 431
Qy	407 -----GDTTYCTE-----NG-W-SPPPKC 424
Dd	432 HTGKPLEVFPFGKAVNYTCDPHPDRGTSTFDLIGESTIRCTSDPCNGCWSSPADRC 487

RESULT 15
IQ16744
ID Q16744 PRELIMINARY; PRT; 2489 AA.
AC Q16744;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE COMPLEMENT RECEPTOR 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN NCBI_TaxID=9606;
RN [1]
RX MEDLINE=94065175; PubMed=8245463;
RA Vik D.P., Wong W.W.;
RT "Structure of the gene for the F allele of complement receptor type 1
RL and sequence of the coding region unique to the S allele.";
RL J. Immunol. 151:6214-6224(1993).
RN [2]
RX SEQUENCE FROM N.A.
RA Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
RL EMBL; L17418; AAB60695.1;
DR EMBL; L17390; AAB60695.1; JOINED.

DR	EMBL; L17399; AAB60695.1; JOINED.
DR	EMBL; L17409; AAB60695.1; JOINED.
DR	EMBL; L17419; AAB60695.1; JOINED.
DR	EMBL; L17420; AAB60695.1; JOINED.
DR	EMBL; L17421; AAB60695.1; JOINED.
DR	EMBL; L17422; AAB60695.1; JOINED.
DR	EMBL; L17423; AAB60695.1; JOINED.
DR	EMBL; L17424; AAB60695.1; JOINED.
DR	EMBL; L17425; AAB60695.1; JOINED.
DR	EMBL; L17426; AAB60695.1; JOINED.
DR	EMBL; L17427; AAB60695.1; JOINED.
DR	EMBL; L17428; AAB60695.1; JOINED.
DR	EMBL; L17429; AAB60695.1; JOINED.
DR	EMBL; L17430; AAB60695.1; JOINED.
DR	EMBL; L17391; AAB60695.1; JOINED.
DR	EMBL; L17392; AAB60695.1; JOINED.
DR	EMBL; L17393; AAB60695.1; JOINED.
DR	EMBL; L17394; AAB60695.1; JOINED.
DR	EMBL; L17395; AAB60695.1; JOINED.
DR	EMBL; L17396; AAB60695.1; JOINED.
DR	EMBL; L17397; AAB60695.1; JOINED.
DR	EMBL; L17398; AAB60695.1; JOINED.
DR	EMBL; L17400; AAB60695.1; JOINED.
DR	EMBL; L17401; AAB60695.1; JOINED.
DR	EMBL; L17402; AAB60695.1; JOINED.
DR	EMBL; L17403; AAB60695.1; JOINED.
DR	EMBL; L17404; AAB60695.1; JOINED.
DR	EMBL; L17405; AAB60695.1; JOINED.
DR	EMBL; L17406; AAB60695.1; JOINED.
DR	EMBL; L17407; AAB60695.1; JOINED.
DR	EMBL; L17408; AAB60695.1; JOINED.
DR	EMBL; L17410; AAB60695.1; JOINED.
DR	EMBL; L17411; AAB60695.1; JOINED.
DR	EMBL; L17412; AAB60695.1; JOINED.
DR	EMBL; L17413; AAB60695.1; JOINED.
DR	EMBL; L17414; AAB60695.1; JOINED.
DR	EMBL; L17415; AAB60695.1; JOINED.
DR	EMBL; L17416; AAB60695.1; JOINED.
DR	EMBL; L17417; AAB60695.1; JOINED.
DR	HSP; P08603; IHFI.
DR	InterPro; IPR001424; SOD_CU_ZN.
DR	InterPro; IPR000436; Sushi_SCR_CCP.
DR	InterPro; IPR000834; Zn_carbopept.
DR	Pfam; PF00084; sushi; 37.
DR	SMART; SM00032; CCP; 37.
DR	PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_3.
DR	PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW	Receptor.
SQ	SEQUENCE 2489 AA: 272846 MW: CEE1B53F2B4FAF6 CRC64;
Query Match 17.8%; Score 437; DB 4; Length 2489;	
Best Local Similarity 27.7%; Pred. No. 4-2e-31;	
Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;	
Qy	27 GTQATYKRCRGYRTLGIIVKVCNGEWPNSPNCRIKRPCGHPDPFGSFRLAVGSEF 86 - - : : : - : : :
Db	66 GTYLNECRPGYSGRPSIICLNKSNVMTGAADR--CRRKSCRNPDPVPNGMVHVG--I 121 - - : : : - : : :
Qy	87 EFGAKVVYTCDGYQLGEIDYRECDADG----WTNDIPICEVVKC-LPVTELENGRIVS 141 : : - : : : : : : :
Db	122 QFGSIKYSTCKTYRLIGS--SSATCIISGDTVIWDNETPIDCRIPCGLPPT-ITNGDFIS 179 : : : : : : : :
Qy	142 GAAPEDQEYFGVOVRPECNSG-----FKIEGQKEMHCSN-----GLWSNEKPQC-VEI 190 : : : : : : : :
Db	180 TNRE---NFHYGSVVYTRCNPGSGGKKRVFLVGEPSIYCTSDDDGVINGSPAPOCLIPN 236 : : : : : : : :
Qy	191 SCLPPRVENG-----DGIILAPYIKENERFYKQGKFVYKERGDVACTG-SGNMPQ-PSC 244 : : : : : : : :
Db	237 KCTPNVNYGILYSDN---RSLSFNWEVFRCOPGFVMKGPRRVKCOALNKWEPELPSC 293 : : : : : : : :
Qy	245 EEWTLTPYPINGIYTPHRITKHRIID-----DEIRYECKNGFYPATRSPVSKTITG-WIP- 298 : : : : : : : : :

Query Match	17.8%	Score 437;	DB 4;	Length 2489;
Best Local Similarity	27.7%;	Pred. No. 4.2e-31;		
Matches	192;	Conservative 63;	Mismatches 149;	Indels 132; Gaps 31;

QY	27	GTQATYKRCRGYRTLGTVKVKNGEWPVNSPRICRKPCHGPDPPFGSFRFLAVCSEF	86
DB	66	GTYNLCRGYSGRPSIICLNKSVMTGADR--CRKSKRNPPDPVNGVMVWIKG--I	121
QY	87	EFKAKVVYTCDEGYQLLGEIDYRECDADG----WTNDIPICEVVKC-LPYTELENGRIVS	141
DB	122	QFGSQIKYSCTKGYRLIGS--SATCIISGDTVIWNETPICDRIPGLPPT-ITNGDFIS	179
QY	142	GAAPDOEYFVGQVVRPECHSG-----FKIEGQKMHCSEN-----GLWSNEHPQC-VEI	190
DB	180	TNRE---NFHYGSVTVYRCNPGSGGRKRVFLVGPEISYICTNSDDVGINSVGAPOCIIPN	236
QY	191	SLCLPPRVENG-----DGIYLPVYKENERFOYKCKQGQVYKERGDAVCTG--SGNPO--PSC	244
DB	237	KCTPPNVENGILYSDN--RSLFSLNEVVEFCQGFVYMKGPRRVKQALNKWEPELPSC	293
QY	245	EEWTCLTPYIPNGIYTPHRIKRID----DEIRYECKNGFYPATRSPVSKCTITG-WIP--	298

Db 294 SRV-COPP--PDVLHAERTORDKDNFSPGOEVFYSCBPG-YDLRGAASMRCTPOGDWSPA 349
Qy 299 APRCSLKPCD--FPQFKHGRLYYEESRRPYPPV--PICKEYSYCDNGFTT--PSQSYWD 352
Db 350 APTCEVAKSCDDFMGQLLNGRV-----LFPVNLQLCAKVDFVCDEGQFLKGSSASY-- 399
Qy 353 YLRCTVNG---WEPEVPCLRQCIFHYVEYGESSYWQRRYIEGOSAKVQCHSGYSLPN-- 406
Db 400 ---CVLAGMESLWNSSVPVCEQ-----IFCPSPVPVTPNGR 431
Qy 407 -----GQDTYYCTE-----NG-W-SPPPKC 424
Db 432 HTGKPLEVPFPGKAVNYTCDPHDPDRGTSFDLIGESTIRCTSDPOGNGVWSSPAPRC 487

Search completed: August 29, 2002, 15:11:44
Job time: 605 sec

GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 15:03:40 ; Search time 75.55 Seconds
 (without alignments)
 629.247 Million cell updates/sec

Title: US-09-316-163-14
 Perfect score: 2454
 Sequence: 1 EDCKGPPRENSILSGWS.....DTYYCTENGWSPPKVCVRK 428

Scoring table:
 BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	990	40.3	240 18 AAW39154	Human partial Comp
2	918	37.4	216 18 AAW39155	Clone PRRB9FH410 C
3	544.5	22.2	578 20 AAY09065	Human complement f
4	481.5	19.6	581 12 AAR13490	Human C4 binding p
5	438	17.8	543 13 AAR28547	CRI-4 (52S, 53S, 5
6	437	17.8	453 20 AAY55757	Human CRI protein
7	437	17.8	778 19 AAW73147	Amino acid sequenc
8	437	17.8	1930 19 AAW45899	Human complement r
9	437	17.8	2039 20 AAY55751	Human C3b/C4b rece
10	437	17.8	2039 22 ABG00287	Novel human diagno
11	437	17.8	2044 22 ABB11782	Human CRI protein

12	437	17.8	2044	22	AAW39224	Human polypeptide
13	437	17.8	2044	22	AAW41010	Human polypeptide
14	437	17.8	2317	10	AAW2219	CRI protein. Homo
15	434	17.7	496	20	AAW55752	Human CRI protein
16	434	17.7	543	13	AAR28557	CRI-4 (99H, 103E)
17	434	17.7	1537	12	AAR11982	Partial human comp
18	434	17.7	1929	22	ABG00103	Novel human diagno
19	434	17.7	2039	14	AAR36743	CRI-4 (266-274 KLK
20	432	17.6	543	13	AAR28570	Human complement t
21	432	17.6	2039	12	AAW11810	Human CRI protein
22	431	17.6	450	20	AAW55754	Human CRI protein
23	430	17.5	543	13	AAR28560	CRI-4 (114S) analo
24	429	17.5	543	13	AAR28548	CRI-4 (57V, 59K) a
25	428	17.4	543	13	AAR28550	CRI-4 (64K) analog
26	428	17.4	543	13	AAR28553	CRI-4 (85R, 87N) a
27	428	17.4	543	13	AAR28558	CRI-4 (109N, 110A,
28	428	17.4	543	13	AAR28565	CRI-4 (121Q) analo
29	428	17.4	543	13	AAR28571	CRI-4 (364-367 NAA
30	427	17.4	543	13	AAR28545	CRI-4 (37Y) analog
31	427	17.4	543	13	AAR28552	CRI-4 (78T, 79D) a
32	426	17.4	543	13	AAR28555	CRI-4 (92T) analog
33	426	17.4	543	13	AAR28563	CRI-4 (117P) analo
34	426	17.4	543	13	AAR28567	CRI-4 (318-321 RNP
35	425	17.3	543	13	AAR28549	CRI-4 (64K, 65T) a
36	425	17.3	543	13	AAR28551	CRI-4 (65T) analog
37	425	17.3	543	13	AAR28562	CRI-4 (116K) analo
38	425	17.3	543	13	AAR28568	CRI-4 (347T, 349Y)
39	424	17.3	543	13	AAR28569	CRI-4 (369-376 STK
40	424	17.3	645	21	AAW53125	Maceca mulatta rha
41	423	17.2	543	13	AAR28556	CRI-4 (94H) analog
42	423	17.2	543	13	AAR28564	CRI-4 (116K, 117P)
43	422	17.2	543	13	AAR28561	CRI-4 (115T) analo
44	421	17.2	543	13	AAR28554	CRI-4 (92T, 94H) a
45	420	17.1	543	13	AAR28544	CRI-4 (35E) analog

ALIGNMENTS

RESULT 1
 AAW39154
 ID AAW39154 standard; Protein; 240 AA.
 XX
 AC AAW39154;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Human partial Complement factor H protein fragment 1.
 DE
 DE Complement factor H; tumour associated antigen; renal cancer;
 KW urogenital cancer; medicament; modulator.
 KW
 OS Homo sapiens.
 XX
 PN WO9738136-A1.
 XX
 PD 16-OCT-1997.
 PD
 XX
 PF 09-APR-1997; 97WO-US05710.
 XX
 PR 06-MAR-1997; 97US-0812481.
 PR 09-APR-1996; 96US-0015083.
 PR 09-APR-1996; 96US-0630048.
 PR 06-MAR-1997; 97US-0038614.
 XX
 PA (BARD-) BARD DIAGNOSTIC SCI INC.
 XX
 PI Enfield DL, Hass GM, Kinders RJ;
 XX
 DR WPI: 1997-512742/47.
 DR N-PSDB; AAV02790.
 XX
 PT Treating or screening for cancer, e.g. renal or urogenital cancer -

PT by modulating or detecting tumour associated human complement Factor
 PT H related antigen, or nucleic acid encoding it
 XX
 PS Example 6B; Fig 6B; 104pp; English.
 XX

CC This partial protein sequence represents a region of the human
 CC tumour-associated complement factor H (CFH). This sequence is used
 CC in the identification of complement factor H related proteins and
 CC antigens isolated from clone PRBB9FH410 (see AAW39155). The detection of
 CC such proteins and a CFH antigens can be used in screening or for the
 CC treatment of renal or urogenital cancer, e.g. bladder, cervical or
 CC prostate cancer. Agents that may modulate this antigen could be used in
 CC the manufacture of a medicament for the treatment of a tumour cell.
 XX
 SQ Sequence 240 AA;

Query Match 40.3%; Score 990; DB 18; Length 240;
 Best Local Similarity 71.7%; Pred. No. 1e-67;
 Matches 172; Conservative 21; Mismatches 47; Indels 0; Gaps 0;

QY 78 FLAVGSEFEFGAKVYTCDEGYQLLGEIDYRECDADGWTNDIPICEVVKCLPVTLENG 137
 DB 1 flltgnvfeygkvavtncnegyllgeinrecdtgwtndipicevkvclpvtapeng 60
 QY 138 RIVSGAAEPDQYYFGQVRFECNSGFKIEGQKEMHCSENGLSNEKPCQCVETISCLPPRV 197
 DB 61 kiwssamepdreyhfgavrfvcnsgyklegdeemhcsddgfwskpkcveiscskspdv 120
 QY 198 ENGSDGYLKPVKYKNERFOYKQGVYKRGDAVCTGSGWNPPQSCSEMTCLTPYIPNG 257
 DB 121 ingdspisqliykenerfykcmgyeysergdavctesgwrplpsceekscdnpyping 180
 QY 258 IYTPHRIKHIDDEIRYECKNGFYPATRSPVSKCTITGWIPAPRCSLKPCDFQFKHGL 317
 DB 181 dysplrikhrtgdeityqrcngfyatrgntaktctgtwipaprcctlkpcdydpdkhgg1 240

RESULT 2
 AAW39155
 ID AAW39155 standard; Protein; 216 AA.
 XX
 AC AAW39155;
 XX

DT 27-APR-1998 (first entry)
 DE
 DE Clone PRBB9FH410 CFH related protein fragment.
 XX
 KW Complement factor H; tumour associated antigen; renal cancer;
 KW urogenital cancer; medicament; modulator.
 XX
 OS Synthetic.
 OS
 PN WO9738136-A1.
 XX
 PD 16-OCT-1997.
 XX

PF 09-APR-1997; 97WO-US05710.
 XX
 PR 06-MAR-1997; 97US-0812481.
 PR 09-APR-1996; 96US-0015083.
 PR 09-APR-1996; 96US-0630048.
 PR 06-MAR-1997; 97US-0038614.
 XX
 PA (BARD-) BARD DIAGNOSTIC SCI INC.
 XX
 PI Enfield DL, Hass GM, Kinders RJ;
 XX
 DR WPI; 1997-512742/47.
 DR N-PSDB; AAW02791.
 XX

PT Treating or screening for cancer, e.g. renal or urogenital cancer -
 PT by modulating or detecting tumour associated human complement Factor

PT H related antigen, or nucleic acid encoding it
 XX
 PS Example 6B; Fig 6B; 104pp; English.
 XX

CC This partial protein is found in clone PRBB9FH410 and represents a
 CC complement factor H related protein with homology to a region of the
 CC human tumour-associated complement factor H (CFH). The detection of this
 CC protein and a CFH antigen can be used in screening or for the treatment
 CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
 CC Agents that may modulate this antigen could be used in the manufacture of
 CC a medicament for the treatment of a tumour cell.
 XX
 SQ Sequence 216 AA;

Query Match 37.4%; Score 918; DB 18; Length 216;
 Best Local Similarity 73.1%; Pred. No. 2.8e-62;
 Matches 158; Conservative 19; Mismatches 39; Indels 0; Gaps 0;

QY 95 TCDEGYQLLGEIDYRECDADGWTNDIPICEVVKCLPVTLENGRIVSGAAEPDQYYFGQ 154
 DB 1 tcnegvqllgeinrecdtgwtndipicevkvclpvtapengkiwssamepdreyhfgq 60
 QY 155 VVRFECNSGFKIEGQKEMHCSENGLSNEKPCQCVETISCLPPRVENGDIYLPVKYKNER 214
 DB 61 avrfvcnsgyklegdeemhcsddgfwgkpkcveiscskspdvingspisqliykener 120
 QY 215 FOYKQGVYKRGDAVCTGSGWNPPQSCSEMTCLTPYIPNGIYTPHRIKHIDDEIRY 274
 DB 121 fgykcmgyeysergdavctesgwrplpsceekscdnpypngdysplrikhrtgdeity 180
 QY 275 ECKNGFYPATRSPVSKCTITGWIPAPRCSLKPCDFP 310
 DB 181 qrcngfyatrgntaktctgtwipaprcctlkpcdydp 216

RESULT 3
 AAY09065
 ID AAY09065 standard; Protein; 578 AA.
 XX
 AC AAY09065;
 XX

DT 06-JUL-1999 (first entry)
 DE
 DE Human complement factor H homolog protein.
 XX
 KW Human complement factor H; immunological mechanism; complement reaction;
 KW gene therapy; immune stimulation; haematopoiesis regulation; chemotactic;
 KW tissue growth activity; anti-inflammatory; tumour inhibition;
 KW secretory signal.
 XX
 OS Homo sapiens.
 OS
 PN WO9918200-A1.
 XX
 PD 15-APR-1999.
 XX

PF 02-OCT-1998; 98WO-JP04448.
 XX
 PR 06-OCT-1997; 97JP-0272837.
 XX
 PA (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENT.
 XX
 PI Kato S, Sekine S;
 XX
 DR WPI; 1999-264019/22.
 DR N-PSDB; AAX34737.
 XX

PT Human proteins with secretory signal sequences and nucleotide
 PT sequences, useful in control of proliferation and differentiation of
 PT cells
 XX

Claims 1; Page 55-58; 71pp; English.

This DNA encodes a protein having homology to human complement factor H, which plays a role in the immunological mechanism involving the complement reaction. The protein can also be used as an antigen for preparing antibodies against the protein. The cDNA can be used as a probe for gene diagnosis and the gene for gene therapy, as well as for large-scale expression of the protein. The protein may also have immune stimulating or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, anti-inflammatory activity, tumour inhibition activity, chemotactic/chemokinetic activity, receptor/ligand activity, etc. The protein is identified by the presence of a hydrophobic N-terminal secretory signal region, knowledge of the protein function is not required, as in e.g. methods of expression cloning.

Sequence 578 AA:

Query Match 22.2%; Score 544.5; DB 20; Length 578;
Best Local Similarity 32.3%; Pred No. 2.9e-33;
Matches 136; Conservative 55; Mismatches 147; Indels 83; Gaps 24;

QY 48 CKNGEWPSPSRICKRKPCHP-----GDPFGSFR-----LAVGSEFFGAKVVTCD E 98
Db 15 cangqev-----kpcdfpeiqhgglvksrlrlyfpaaagsys-----yycdq 58
QY 99 GYQL-LGEI-DYRECDADGWINDIPCEVVKLPV-----TELENGRIVSGAAEPQ EY 151
Db 59 nvtptsgsywdyihctqdgwstvp-----clrtcsksadvelengfi-----sesssiyi 108
QY 152 FGQVVRFECSGFKI---EGQEHMCSENGLSNEKPOCWEISCLPPRVEN---GDGIY 204
Db 109 lneetqncpkyatadgnssgsitclqng-wstq-picikf-cdmpvfensraksngmw 165
QY 205 LKPVYKENERFOYKQKGF--VYKERGDA-VCTGSGWNPQPSG---EEMTCLTPYIPNG- 257
Db 166 ----fklhdtldyecdgyessyngntdsivcgedgwhlptcynssescgppppisngd 221
QY 258 -----IYTPHRIKHRIIDETREYCKNGFYATPSVSKCITGWIPAPRC-SLUKPCDF 309
Db 222 ttfspqkylpw-----srveydcqs-yyelqgskyyvtcsngdwsepprcismkpcsf 273
QY 310 POFKIGRLYYEESRRPYFPVPIGKESYYCDNGFTTPSQSYWDYLKRCVTNGWEPEVPCLR 369
Db 274 peiqhghlyventrpyfpvatgqsyysydcndfnvtpsgsywdyihctqdgwlpvtvpcir 333
QY 370 QCIFH--VVEYGESSYWMORRYIEGOSAKVQCHSGYSLPNQGD--TYCTENGWSPPPKCV 425
Db 334 tcsksadieiengfisesessiyilnkeiykckpgyatadgnssgsitclqngwsaqpic 393
QY 426 R 426
Db 394 k 394

RESULT 4

ID AAR13490 standard; Protein; 581 AA.

AC AAR13490;

DT 30-OCT-1991 (first entry)

DE Human C4 binding protein.

KW C4bp; monomer; complement protein; pJOD.C4bp.3; SCR;
short consensus repeat.

OS Homo sapiens.

FT Key Location/Qualifiers
Peptide 1..32

FT Protein /label= signal_peptide
FT 33..581
FT /label= C4bp
FT 33..93
FT /label= SCR8
FT 94..155
FT /label= SCR7
FT 156..219
FT /label= SCR6
FT 220..279
FT /label= SCR5
FT 280..345
FT /label= SCR4
FT 346..406
FT /label= SCR3
FT 407..464
FT /label= SCR2
FT 465..523
FT /label= SCR1
FT 524..581
FT /label= C4bp_core
FT /note= "responsible for multimer assembly"
FT Disulfide-bond 34..80
FT /note= "intradomain"
FT Disulfide-bond 65..92
FT /note= "intradomain"
XX W09111461-A.
PN 08-AUG-1991.
XX 28-JAN-1991; 91WO-US00567.
XX 26-JAN-1990; 90US-0470888.
XX (BIOG-) BIOGEN INC.
PI Pasek MP, Winkler G, Liu TR;
XX WPI; 1991-252613/34.
DR N-PSDB; AAK13242.
XX
XX New C4 binding protein fusion proteins and DNA encoding them -
PT comprise assemblies of C4bp monomers linked to functional moiety,
PT e.g. AZT, useful as delivery vehicles in diagnosis and therapy
XX
PS Example 1; Fig 1; 105pp; English.
XX
XX This sequence was deduced from human hepatocyte (Hep G2) cDNA
CC obtained following PCR amplification. The protein is a monomer
CC containing 8 SCRs. Each SCR forms a looped domain due to the
CC presence of two intradomain disulphide bonds (only the disulphide
CC bonds of SCR8 are labelled in the Features Table). Within each SCR,
CC the first cysteine residue bonds with the third and the second
CC cysteine residue bonds with the fourth. This secondary structure is
CC responsible for the conformational flexibility of the C4bp monomer.
CC The invention covers fusion proteins in which the monomer sequence,
CC or a specified subfragment of it, e.g. having 5, 4, 3 or 1 SCR(s)
CC is fused to the C-terminal of a protein such as a viral receptor,
CC cell ligand, a bacterial, viral or parasitic immunogen, enzyme,
CC cytokine, toxin, etc. See also AAK13243-51.
XX
SQ Sequence 581 AA;

Query Match 19.6%; Score 481.5; DB 12; Length 581;
Best Local Similarity 26.5%; Pred. No. 1.8e-28;
Matches 134; Conservative 78; Mismatches 195; Indels 99; Gaps 25;
QY 2 DCKGPPPRENSEILSGWSEQLYSEGTOATYKCRPGY-RTLGTIVKVC-KNGEWPSPNS 59
Db 33 ncgppptlsfaapmditltetrfktgttktyclpgyvrshstqtlctnsgdewynt--- 90

DE	XX		Human CR1 protein LHR-C fragment.
KW	XX	C3B/C4B receptor; CR1 protein; cell-surface protein; erythrocyte; human;	
KW	XX	complement regulatory activity; complement pathway enzyme; tissue damage;	
KW	XX	reperfusion injury; Arthus reaction; myocardial infarct; inflammation;	
KW	XX	heart condition; autoimmune disorder; long homologous repeat; LHR.	
OS	XX	Homo sapiens.	
XX	XX	U55981481-A.	
PN	XX	09-NOV-1999.	
PD	XX	06-JUN-1995; 95US-0470652.	
PF	XX	03-APR-1989; 89US-0332865.	
PR	XX	06-DEC-1974; 74US-0350238.	
PR	XX	24-FEB-1993; 93US-0026134.	
PR	XX	01-APR-1988; 88US-0176532.	
PA	XX	(UYJO) UNIV JOHNS HOPKINS.	
PA	XX	(BGHM) BRIGHAM & WOMENS HOSPITAL.	
PA	XX	(AVAN-) AVANT IMMUNOTHERAPEUTICS INC.	
XX	XX	Concino MF, Wong WW, Makrides SC, Klickstein LB, Fearon DT, Ip SH;	
PI	XX	Marsh HC, Carson GR;	
PT	XX	WPI; 1999-633357/54.	
DR	XX	A human C3B/C4B receptor (CR1) protein having antiinflammatory and	
PT	XX	cardiant activity -	
XX	XX	Disclosure; Fig 5B; 87pp; English.	
PS	XX	The invention relates to a human C3B/C4B receptor (CR1) protein. The CR1	
XX	XX	protein or fragment is expressed as a cell-surface protein on the surface	
CC	XX	of a non-human cell and exhibits a complement regulatory activity of full	
CC	XX	-length human CR1 as expressed on erythrocytes. The CR1 function in vivo	
CC	XX	may be mediated through the inhibition of complement pathway enzymes. The	
CC	XX	soluble CR1 protein exhibits a complement regulatory activity, and this	
CC	XX	may be used to prevent reperfusion injury, inhibit Arthus reaction, and	
CC	XX	neutrophil mediated tissue damage, and reduce myocardial infarct size,	
CC	XX	and inflammation. The CR1 protein and its fragments can also be used in	
CC	XX	the treatment of conditions which involve unwanted complement activity,	
CC	XX	e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,	
CC	XX	and autoimmune disorders. CR1 proteins, analogues, derivatives, and anti	
CC	XX	-CR1 antibodies are used in assays, and diagnostics. The present sequence	
CC	XX	represents the human CR1 protein long homologous repeat (LHR)-C fragment.	
XX	XX	Sequence 453 AA;	
SQ	XX		
Query Match		17.8%; Score 437; DB 20; Length 453;	
Best Local Similarity		29.6%; Pred. No. 3.3e-25;	
Matches 136; Conservative		64; Mismatches 161; Indels 98; Gaps 32;	
QY	27	GTOATYKRCRGYRTLTIVKVCNKEWPSPNRICRKPCGHPGDTPGCSFLVAGSEF 86	
DB	26	gtslkycrpeyygrfptsitcdlnlw--sspdvckrkscktppdpvng--mvhvitdi 81	
QY	87	EFGAKVVYTCEGYQLIGEDIDRECDADG---WTNDIPICEVVKC-LPVTLENGRIVS 141	
DB	82	qvsgrinsycttghrligh-saacillsngntahwstktpicripqlppt-langdfis 139	
QY	142	GAAREPQEYYTGVGFVEFCNSG-----PKIEGQKEMHCSEN-----GLMSNEKPQC-VET 190	
DB	140	tnre----nfhygsavvtircnlgrrkrkvelvgepslyctsnddqvgiwsgpapqiipn 196	
QY	191	SCUPPVENG-----DGILKLPPYKENEREQYKCKQGCVVKERCDAVCTG-SGNPQP-PSC 244	
DB	197	kctppnvengilvsdn---rsifslnevvefcqqpfvmkgprvrvcckqalnkwepelpsc 253	
QY	245	EEWTCLTYPINGIYTP-HRIKHRIIDEIRYCKNGFYPATRSPVSKTCITG-WIP-APR 301	

Query Match 17.8%; Score 437; DB 19; Length 778;
 Best Local Similarity 27.7%; Pred. No. 6.6e-25;
 Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;

PT Derivatives of soluble poly-peptide(s) bonded to low affinity
 PT membrane binding groups - useful for treating complement-related and
 PT thrombotic diseases, providing improved localisation at cellular
 PT membranes
 XX Claim 22; Pages 60-61; 75pp; English.
 XX This sequence represents human complement receptor 1 (CRI, CD 35)
 CC N-terminal fragment. The invention relates to a soluble derivative (A)
 CC of a soluble polypeptide (I), which comprises at least 2 heterologous
 CC membrane-binding elements (MBE) of low membrane affinity covalently
 CC associated with (I). MBE interact, independently and with thermodynamic
 CC additivity, with components of cellular or artificial membranes exposed
 CC to extracellular fluids. (A) are used to treat disorders treatable with
 CC (I) itself, specifically inflammation or any other complement-related
 CC disorder (e.g. neurological disease, graft rejection, myocardial
 CC infarction, sepsis, rheumatoid arthritis and many others; including
 CC application to indwelling devices) and thrombolytic disease, but also to
 CC treat allergy, induce weight loss, to treat ischaemia or asthma and as
 CC immuno-modulators for treating multiple sclerosis. (A) are administered
 CC orally, topically, by injection or inhalation at 0.01-10 (preferably
 CC 0.1-10) mg/kg/day.
 XX Sequence 1930 AA;

QY 27 GTQATYKCRPGYRTLTGTVKCKNGEWPSPNSRICKRKPCHPGDTPFGSFLAVGSEF 86
 Db 71 gylneyecrpgysgrpfsliclknsvwtgakt--crkscrnppdpvngmhviki--1 126
 QY 87 EFGAKVVTCDGEGYQLLGBIDYRECDAG----WTNDIPICEVWK-LPVTLENGRIVS 141
 Db 127 qfgsqikysctkgrylgs-saatclisgdtvwdnetpicdripcpt-itngdfis 184
 QY 142 GAAEPDQYFGQVVRFECSG-----FKIEGQKEMHCSEN-----GLWSNEKPOC-V 190
 Db 185 ture---nfhygsvvtyrcnpgsggrkvfelvgepsiyctsnddqvlgswgapqciipn 241
 QY 191 SCLPPRVENG-----DGIYLPVYKENERFOYKCKOGFYVYKERGDVCTG-SGWNPO-PSC 244
 Db 242 kctppnvengilvsdn---rsifsnvvefrccqpfvmkgprvrkqcalnkweelpsc 298
 QY 245 EEMTCLTPYIPNGIYTPHRIKHID----DEIRYECKNGFYPATRSPVSKCTITG-WIP- 298
 Db 299 srv-cqpp--pdvlhaertqrkdnfspgqevfyscepg-ydlrgaasmrctpggdwsa 354
 QY 299 APRCSLKPCD--FPQFKHGRLYEESRRPYFPV--PICKEYSYCDNGFTT--PSQSYWD 352
 Db 355 aptcevkscddfmqgllngrv-----lfpvnlqlgakvdfvdegfqlkgssasy-- 404
 QY 353 YLRCTVNG-----WEPEVFLQRCIFHYVEYGESSYQWRRYIEGQSAKVQCHSGYSLPN-- 406
 Db 405 ---cvlagmeslwnssvpvceq-----GQDTYYCTE-----NG-W-SPPPKC 424
 QY 437 htgkplevfpfgkavnytcdbphdrgtsfdlgestirctsdpgngvwsaparc 492

RESULT 8
 AAW45899 standard; peptide; 1930 AA.
 XX AAW45899;
 AC AAW45899;
 DT 30-JUN-1998 (first entry)
 XX Human complement receptor 1 (residues 1-1929).
 DE Membrane binding element; thrombotic disease; soluble protein;
 KW complement-related disease; integral membrane protein; inflammation;
 KW short consensus repeat; SCR 1-3; CRI; complement receptor type 1.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FT Cross-links 1930
 FT /note= "Disulphide linked to Cys in peptide given
 FT in AAW45899"
 XX WO9802454-A2.
 PN 22-JAN-1998.
 XX 08-JUL-1997; 97WO-EP03715.
 XX 15-JUL-1996; 96GB-0014871.
 PR (ADPR-) ADPROTECH PLC.
 XX Dodd I, Mossakowska DEI, Smith RAG;
 XX WPI; 1998-110524/10.
 DR
 XX

QY 27 GTQATYKCRPGYRTLTGTVKCKNGEWPSPNSRICKRKPCHPGDTPFGSFLAVGSEF 86
 Db 71 gylneyecrpgysgrpfsliclknsvwtgakt--crkscrnppdpvngmhviki--1 126
 QY 87 EFGAKVVTCDGEGYQLLGBIDYRECDAG----WTNDIPICEVWK-LPVTLENGRIVS 141
 Db 127 qfgsqikysctkgrylgs-saatclisgdtvwdnetpicdripcpt-itngdfis 184
 QY 142 GAAEPDQYFGQVVRFECSG-----FKIEGQKEMHCSEN-----GLWSNEKPOC-V 190
 Db 185 ture---nfhygsvvtyrcnpgsggrkvfelvgepsiyctsnddqvlgswgapqciipn 241
 QY 191 SCLPPRVENG-----DGIYLPVYKENERFOYKCKOGFYVYKERGDVCTG-SGWNPO-PSC 244
 Db 242 kctppnvengilvsdn---rsifsnvvefrccqpfvmkgprvrkqcalnkweelpsc 298
 QY 245 EEMTCLTPYIPNGIYTPHRIKHID----DEIRYECKNGFYPATRSPVSKCTITG-WIP- 298
 Db 299 srv-cqpp--pdvlhaertqrkdnfspgqevfyscepg-ydlrgaasmrctpggdwsa 354
 QY 299 APRCSLKPCD--FPQFKHGRLYEESRRPYFPV--PICKEYSYCDNGFTT--PSQSYWD 352
 Db 355 aptcevkscddfmqgllngrv-----lfpvnlqlgakvdfvdegfqlkgssasy-- 404
 QY 353 YLRCTVNG-----WEPEVFLQRCIFHYVEYGESSYQWRRYIEGQSAKVQCHSGYSLPN-- 406
 Db 405 ---cvlagmeslwnssvpvceq-----GQDTYYCTE-----NG-W-SPPPKC 424
 QY 437 htgkplevfpfgkavnytcdbphdrgtsfdlgestirctsdpgngvwsaparc 492

RESULT 9
 AAY55751
 ID AAY55751 standard; Protein; 2039 AA.
 XX AAY55751;
 AC AAY55751;
 DT 22-FEB-2000 (first entry)
 XX Human C3b/C4b receptor (CRI) protein.
 DE
 XX

C3B/C4B receptor; CRI protein; cell-surface protein; erythrocyte; human; complement regulatory activity; complement pathway enzyme; tissue damage; reperfusion injury; Arthus reaction; myocardial infarct; inflammation; heart condition; autoimmune disorder; diagnostic.

Homo sapiens.

US5981481-A

09-NOV-1999

06 - TUN - 1005 - 0500 - 0470652

[illegible]

06-DEC-1974; 74US-0350238.

24 FEB 1993; 3503 00Z0134;
01-APR-1988; 88US-0176532.

(UYJO) UNIV JOHNS HOPKINS.

(BGHM) BRIGHAM & WOMENS HOSPITAL
(AVAN-) AVANT IMMUNOTHERAPY

Concise ME	Wong MW	Matrillas SO
------------	---------	--------------

Marsh HC, Carson GR;

WPI; 1999-633357/54.

ACCTOCTU, 0001 A

A human C3B/C4B Fc
cardiant activity

Disclosure: FIA 1A-P. 8700. English

044
 045
 046
 047
 048
 049
 050
 051
 052
 053
 054
 055
 056
 057
 058
 059
 060
 061
 062
 063
 064
 065
 066
 067
 068
 069
 070
 071
 072
 073
 074
 075
 076
 077
 078
 079
 080
 081
 082
 083
 084
 085
 086
 087
 088
 089
 090
 091
 092
 093
 094
 095
 096
 097
 098
 099
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555

protein or fragment is expressed as a cell-surface protein on the surface of a non-human cell and exhibits a complement regulatory activity of full-length human CR1 as expressed on erythrocytes. The CR1 function *in vivo* may be mediated through the inhibition of complement pathway enzymes. The soluble CR1 protein exhibits a complement regulatory activity, and this may be used to prevent reperfusion injury, inhibit Arthus reaction, and neutrophil mediated tissue damage, and reduce myocardial infarct size, and inflammation. The CR1 protein and its fragments can also be used in the treatment of conditions which involve unwanted complement activity, e.g. shock lung, tissue damage due to burn, or ischemic heart conditions, and autoimmune disorders. CR1 proteins, analogues, derivatives, and anti-CR1 antibodies are used in assays, and diagnostics. The present sequence represents the human CR1 protein.

Sequence 2039 AA:

Query Match 17.8%; Score 437; DB 20; Length 2039;
Best Local Similarity 27.7%; Pred. NO. 2.2e-24;
Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;

```

QY      27   GTOATYKCRPGYRTLGTIVKVKNGEWPNSPRICRRKPCGHPGDTPFGSRLAVGSEF    86
        || |.|.||||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      66   gtylnycrpgysgrpsfiiclknsvgtgakt--crlkscrnpdpbvngmnhvikg--i 121

QY      87   EFGAKVVTYDCEGQLLGEDIDYRECDADG-----WTNDIPICEVVKC-LPVTELENGRIVS 141
        :||:: :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db     122   qfsgqikysctkgrylligs-ssatciisgdvtivdwnetpicdrpcglppt-ltngdfis 179

QY     142   GAAPDPQEYYFGOVVRFECSNG-----FKLEGOKEMHCSEN-----GLWSNEKFQC-VEI 190
        :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db     180   tnre---nfygsvvtyrcnpgsggrkvfelvgpeislcytsnddqvgvlswapqcqlpn 236

QY     191   SCLPPRVENG-----DGILYLKPVKYENERFOYKQGQGVYKERGDVACTG-SGMNPQ-PSC 244
        :|| |.||||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db     237   kctppnvengilvsdn--tslslnevfvefcqpfvmkgyprrvkcaqlnkwepeipsc 293

QY     245   EEMTCLTPYPNGIYTPHRTKHRIID-----DEIRYECKNGFYPATRSPVSXKTITG-WIP- 298
        :| |:| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```


QY 191 SCLPRVENG-----DGIYLPVKYKENERFOYKCKOGFVYKRGDAVCTG--SQWNPQ--PSC 244
 Db 242 kctpnvengilvsdn---rsflsinevtrcpgfvmkgprrrvkcaqlnkweelpsc 298
 QY 245 EEMTCLTPYIPNGIYTPHRIKHRIID-----DEIRYCKNGFYFATPSVSKCTITG--WIP- 298
 Db 299 srv-cqpp--pdvIhaertqrdknfsgpgqevfyscepg-ydlrgaasmrctpgqdwspa 354
 QY 299 APRCSLKPCD--FPQFKHGRLYEESRRPYFPV--PICKEYSYVCDNGFTT--PSQSYWD 352
 Db 355 aptcevkscddfmqllngrv-----lfpvnqlgkdvfcvcdgfkkgssasy-- 404
 QY 353 YLRCTVNG-----WEPEVPCLRQCIFHYVYEGESSYQWRRYIEGQSAKVOCHSGYSLPN-- 406
 Db 405 ---cvIagmeslwnssvpvceq-----ifcspvpvpingr 436
 QY 407 -----GQDTYYCTE-----NG-W-SPPPKC 424
 Db 437 htgkplevfpfgkavnytcdbphdrgtsfdligestirctsdpgqngvwwspaprc 492

RESULT 12
 AAM39224
 ID AAM39224 standard; Protein; 2044 AA.
 XX
 AC AAM39224;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 2369.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AAI58380.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 4; SEQ ID NO 2369; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 2044 AA;
 Query Match 17.8%; Score 437; DB 22; Length 2044;
 Best Local Similarity 27.7%; Pred. No. 2.2e-24;
 Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;
 QY 27 GTQATYKCRPGYRTLTIVKVKNGEWPSPNPSRICRKRPCGHPGDTFGSFLAVGSEF 86
 Db 71 gtylnecrpgysgrpfsiclknsvwgakdr--crrksrnpdpvngmrvhikg--i 126
 QY 87 EFGAKVYTCDEGYQLLGEIDYRECDAG-----WTNDIPICEVVKC-LPVTLENGRIVS 141
 Db 127 qfsgikysctkgyrligs--ssatcilsqdtvwdnetpdrpcglprrt-itngdfis 184
 QY 142 GAAEPDQYVFGQVYRFECNSG-----FKTEGQKEMHCSEN-----GLWSNEKPCQ-V 190
 Db 185 tnre---nfhygsvvtvrcnpgsggrkvfelvgpslyctsdnddqvlgwsgpapcqlpn 241
 QY 191 SCLPRVENG-----DGIYLPVKYKENERFOYKCKOGFVYKRGDAVCTG--SQWNPQ--PSC 244
 Db 242 kctpnvengilvsdn---rsflsinevtrcpgfvmkgprrrvkcaqlnkweelpsc 298
 QY 245 EEMTCLTPYIPNGIYTPHRIKHRIID-----DEIRYCKNGFYFATPSVSKCTITG--WIP- 298
 Db 299 srv-cqpp--pdvIhaertqrdknfsgpgqevfyscepg-ydlrgaasmrctpgqdwspa 354
 QY 299 APRCSLKPCD--FPQFKHGRLYEESRRPYFPV--PICKEYSYVCDNGFTT--PSQSYWD 352
 Db 355 aptcevkscddfmqllngrv-----lfpvnqlgkdvfcvcdgfkkgssasy-- 404
 QY 353 YLRCTVNG-----WEPEVPCLRQCIFHYVYEGESSYQWRRYIEGQSAKVOCHSGYSLPN-- 406
 Db 405 ---cvIagmeslwnssvpvceq-----ifcspvpvpingr 436
 QY 407 -----GQDTYYCTE-----NG-W-SPPPKC 424
 Db 437 htgkplevfpfgkavnytcdbphdrgtsfdligestirctsdpgqngvwwspaprc 492

RESULT 13
 AAM41010
 ID AAM41010 standard; Protein; 2044 AA.
 XX
 AC AAM41010;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 5941.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX

PN WO200153312-A1.
XX 26-JUL-2001.
PD 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDB; AAI60166.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Example 2; SEQ ID NO 5941; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAI38642-AAI42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX Sequence 2044 AA;

Query Match 17.8%; Score 437; DB 22; Length 2044;
Best Local Similarity 27.7%; Pred. No. 2.2e-24;
Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;
QY 27 GTOATYKRCPCGYRTLGTITVCKNGEWPVNSPRTICRRKPCGHPDTPFGSFLAVGSEF 86
DB 71 gtylnyrcpysgrpfisiclknsvtgakdr--crrksrnpdpvgmvhvkig--i 126
QY 87 EFGAKVYVTCDEGQLLGEIDYRECDADG----WTNDIPICEVVKC-LPVTLENGRIVS 141
DB 127 qfsgikysctkgryllgs--ssatciilsgdvtviwdnetpicdrpcglpbt-ingdfis 184
QY 142 GAAPDEQYFYGQVVRRECNSG-----FKIEGQKEMHCSEN----GLWSNEKPOC-VEI 190
DB 185 tnre---nfhygsvvtvrcpnsgrkrvfeivgepsiyctsnddqvgiwsqapqciipn 241
QY 191 SCLPPRVENG----DGIYLVKPYKRENERFOYKCKQGQGVYKERGDVCTG-SGWNPO-PS 244
DB 242 kctppnvengilvsdn---rsfslnevfeircpgvmkgprvrkcaqlnkweipsc 298
QY 245 EEMTCLTPYIPNGIYTPHRIKHRID----DEIRYECKNGFYPAATRSVPVSKCTITG-WLP- 298
DB 299 srv-cqpo--pdvllaertqrdkdnfsgpgqevfyscepg-ydlrgaasmrctpggdwsa 354

QY 299 APRCSLKPCD--FPQFKHGRLLYYEESRRPYFPV--PIGKEYSYCDNGFTT--PSOSYWD 352
DB 355 aptcevkscddfmqqlngrv-----lfpvniqlgakvdfvcdgqlkgssasy-- 404
QY 353 YLRCTVNG-----WEPEVCLROCIEFHYVEYGESSYWORRYIEGQSAKVOCHSGYSLPN-- 406
DB 405 ---cvlagmeslwnssvpvceq-----ifcspppvlpngr 436
QY 407 -----GQDTYYCTE-----NG-W-SPPKRC 424
DB 437 htgkplevfpfgkavnyctdphdrgtsfdligestirctsdpgngvswsspaprc 492
RESULT 14
AAP92219
ID AAP92219 standard; protein; 2317 AA.
XX AC AAP92219;
XX AC AAP92219;
DT 22-FEB-1990 (first entry)
XX CR1 protein.
DE Complement; cofactor.
KW Homo sapiens (human).
OS Homo sapiens (human).
FH Key Location/Qualifiers
FT Peptide 10..50
FT /label= signal_peptide
PN WO8909220-A.
XX 05-OCT-1989.
PD 31-MAR-1989; 89WO-US01358.
XX 01-APR-1988; 88US-0176532.
XX (TCEL) T CELL SCIENCES INC.
PA (UWJO) THE JOHNS HOPKINS UNIVERSITY.
PA (BRIG) THE BRIGHAM AND WOMEN'S HOSPITAL.
XX Fearon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;
XX WPI: 1989-309498/42.
XX N-PSDB; AAN91477.
XX New nucleic acid sequences encoding new CR1 protein - and its fragment,
PT for diagnosis and control of complement-related immune defects,
PT inflammation, myocardial infarct, etc
XX Claim 1; Fig 1; 191pp; English.
XX This is full-length CR1 protein, and shortened forms are new, lacking
CC the transmembrane region. The proteins and fragments bind C3b and/or
CC C4b, have cofactor activity and inhibit C3 and C5 convertase activity.
CC In the sequence, x-untranslated region. This has 7 short consensus
CC repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in
CC patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C
CC for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.
CC They are useful in diagnosing and treating immune disorders, and prevent
CC perfusion injury.
XX Sequence 2317 AA;
SQ
Query Match 17.8%; Score 437; DB 10; Length 2317;
Best Local Similarity 27.7%; Pred. No. 2.6e-24;
Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;
QY 27 GTOATYKRCPCGYRTLGTITVCKNGEWPVNSPRTICRRKPCGHPDTPFGSFLAVGSEF 86
DB 71 gtylnyrcpysgrpfisiclknsvtgakdr--crrksrnpdpvgmvhvkig--i 126
QY 87 EFGAKVYVTCDEGQLLGEIDYRECDADG----WTNDIPICEVVKC-LPVTLENGRIVS 141
DB 127 qfsgikysctkgryllgs--ssatciilsgdvtviwdnetpicdrpcglpbt-ingdfis 184
QY 142 GAAPDEQYFYGQVVRRECNSG-----FKIEGQKEMHCSEN----GLWSNEKPOC-VEI 190
DB 185 tnre---nfhygsvvtvrcpnsgrkrvfeivgepsiyctsnddqvgiwsqapqciipn 241
QY 191 SCLPPRVENG----DGIYLVKPYKRENERFOYKCKQGQGVYKERGDVCTG-SGWNPO-PS 244
DB 242 kctppnvengilvsdn---rsfslnevfeircpgvmkgprvrkcaqlnkweipsc 298
QY 245 EEMTCLTPYIPNGIYTPHRIKHRID----DEIRYECKNGFYPAATRSVPVSKCTITG-WLP- 298
DB 299 srv-cqpo--pdvllaertqrdkdnfsgpgqevfyscepg-ydlrgaasmrctpggdwsa 354

